Laboratorio N.3

Introduccion a Los Metodos Estadisticos
Estimacion por intervalos y simulacion

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

Universidad Del Valle
Facultad De Ingenieria
Estadistica
Octubre
2017

Índice

1. Situación 1

Se generaron 5000 muestras aleatorias de tamaño n=10, de una población normal con parámetros $\mu = 5$ y $\sigma = 1$ y para cada una de las muestras, se encontró un intervalo de confianza para la media y la varianza respectivamente, con una confianza del 95 %.

Para la media:

Nos arrojo que el porcentaje de intervalos que atrapan la verdadera media es: 95.22 %

Y la longitud promedio o esperada de cada intervalo fue hallada como: $\sum_{i=1}^{5000} \frac{(LS_i - LI_i)}{5000}$ lo que nos arrojo un resultado de 1.2395.

El porcentaje de intervalos que atrapan la verdadera media tiene mucho sentido, ya que, al trabajar con una confianza del 95 %, estamos diciendo que el 95 % de las veces que se repita el experimento (en este caso que se obtenga una muestra distinta de la misma población), el verdadero valor de la media caerá en el intervalo obtenido. Como obtuvimos un porcentaje del 95.22 %, nos indica que el 95.22 % de las veces que se repitió el experimento (remuestrear y obtener el IC para la media), la verdadera media cayo en el intervalo encontrado.

Con respecto a la longitud promedio, esta nos arroja la amplitud que tiene aproximadamente cada intervalo. En este caso, todos los intervalos tienen exactamente la misma longitud, ya que utilizamos la estimación cuando σ es conocida, y en este tipo de estimación, la parte que se le suma y se le resta a la media muestral para hallar el intervalo de confianza, no depende de los datos. esa parte es $\pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. Entonces,en este caso, el 1.2395 que nos arrojo la longitud promedio,nos dice que cada uno de los 5000 intervalos tiene una longitud de 1.2395.

Para la varianza:

1.1. Punto a.

PARA LA MEDIA:

Utilizamos la siguiente tabla para ver mas fácilmente la comparación por tamaños de muestra, de la proporción de intervalos que atrapan la verdadera media y de la longitud promedio de los intervalos.

	n=10	n=30	n=50	n=100
Porcentaje de cubrimiento	0.9522	0.9474	0.9508	0.9464
Longitud promedio del intervalo	1.2395	0.7156	0.5543	0.3919

La siguiente figura nos muestra los 100 primeros intervalos con cada tamaño de muestra (10,20,50,100), ya que si elaborábamos la gráfica con los 5000 intervalos estimados, no nos daba una visión clara ni comparable de lo que ocurre.

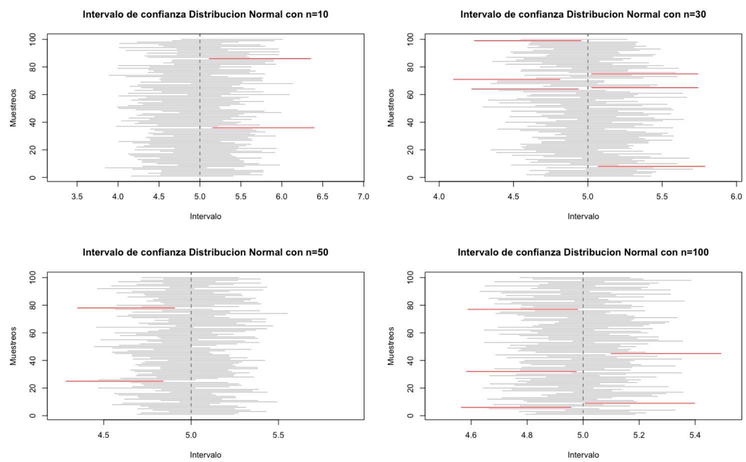

Figura 1: Gráfica comparativa de los 100 primeros intervalos por cada tamaño de muestra

Representación gráfica del porcentaje de cubrimiento esperado de los intervalos de la media:

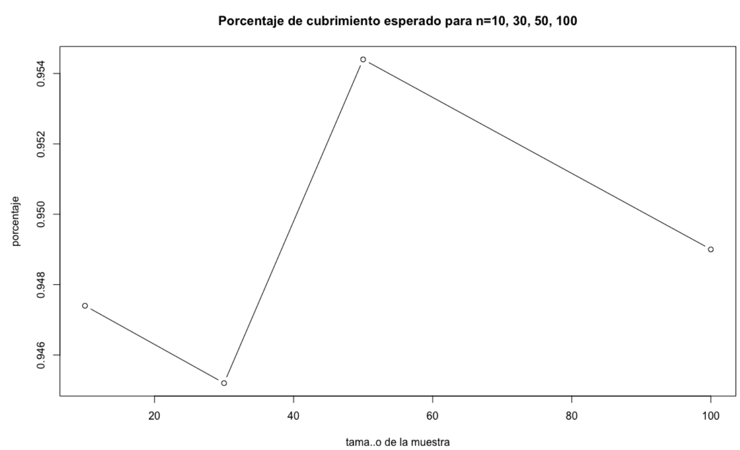

Figura 2: Gráfica del porcentaje de cubrimiento de los intervalos para μ por cada tamaño de muestra

En esta imagen podemos ver que todos los porcentajes de cubrimiento están cerca del 95 %, lo cual es lógico, ya que estamos trabajando con una confianza del 95 %, lo que nos dice que el 95 % de las veces que se repita el proceso (remuestrear y hallar el intervalo de confianza) la verdadera media o la media real, va a caer dentro del intervalo estimado. Ademas podemos ver que el porcentaje de cubrimiento disminuye o aumenta sin depender de el tamaño de muestra n(10,30,50,100), es decir, no se ve un patrón claro de dependencia entre el tamaño de las muestras y el porcentaje de cubrimiento de los intervalos estimados.

Representación gráfica de la longitud esperada por intervalo para la media:

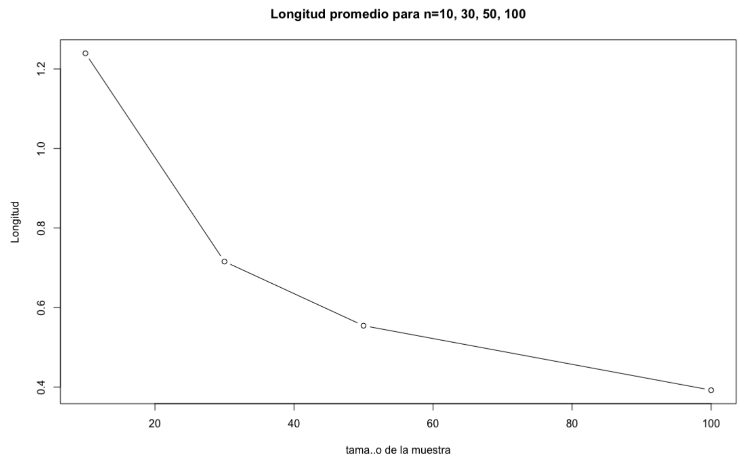

Figura 3: Gráfica de la longitud esperada de cada intervalo para μ por cada tamaño de muestra

En la anterior imagen observamos que la longitud mas alta es la de n=10, y no sobrepasa de 1.5, y la mas baja es la de n=100 y es de aproximadamente 0.4. Esto nos muestra que hay una correlación lineal negativa entre el tamaño de muestra y la longitud del intervalo, es decir, mientras mayor es el tamaño de muestra n, menor es la longitud del intervalo. Lo anterior tiene mucho sentido, ya que la parte que se le suma y se le resta a la media de cada muestra, para hallar una estimación por intervalos para la media real μ es, $Z_{(1-\alpha)} \cdot \frac{\sigma}{\sqrt{n}}$ y vemos que esta expresión es mas pequeña cuando el n aumenta, haciendo menor, la amplitud de cada intervalo.

PARA LA VARIANZA:

Utilizamos la siguiente tabla para ver mas fácilmente la comparación por tamaños de muestra, de la proporción de intervalos que atrapan la verdadera varianza y de la longitud promedio de los intervalos.

	n=10	n=30	n=50	n=100
Porcentaje de cubrimiento	0.9524	0.9454	0.9496	0.9502
Longitud promedio del intervalo	2.8446	1.1712	0.8543	0.5795

Representación gráfica del porcentaje de cubrimiento esperado de los intervalos de la varianza:

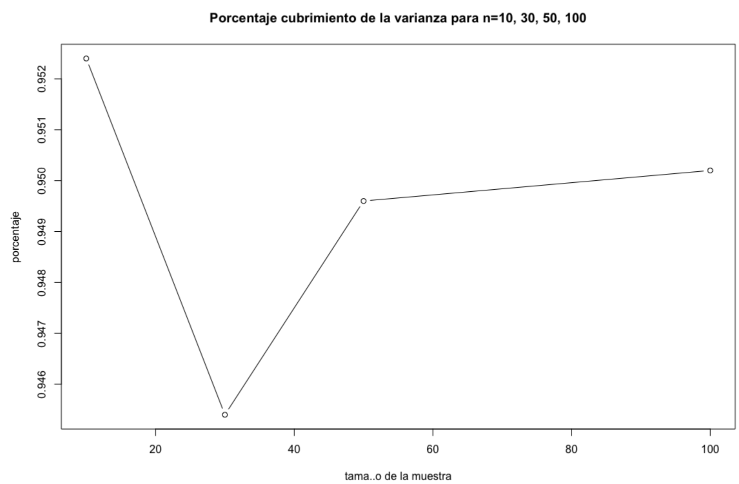

Figura 4: Gráfica del porcentaje de cubrimiento de los intervalos para σ^2 por cada tamaño de muestra

En esta imagen podemos ver que todos los porcentajes de cubrimiento de los intervalos para la varianza están cerca del 95 %, lo cual es lógico, ya que como se explico en el punto anterior, estamos trabajando con una confianza del 95 %, lo que nos dice que el 95 % de las veces que se repita el proceso (remuestrear y hallar el intervalo de confianza) la verdadera varianza o la varianza real, va a caer dentro del intervalo estimado. Ademas podemos ver que el porcentaje de cubrimiento disminuye o aumenta sin depender de el tamaño de muestra n(10,30,50,100), es decir, no se ve un patrón claro de dependencia entre el tamaño de las muestras y el porcentaje de cubrimiento de los intervalos estimados para la varianza.

Representación gráfica de la longitud esperada por intervalo para la varianza:

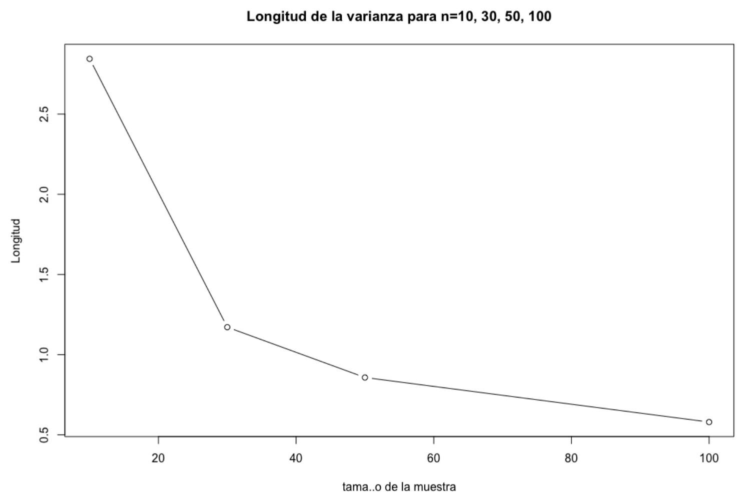

Figura 5: Gráfica de la longitud esperada de cada intervalo para σ^2 por cada tamaño de muestra

En esta imagen, al igual que en la de los intervalos para la media. Vemos que la longitud es mayor para los tamaños de muestra de n=10 y es menor para los tamaños de muestra de n=100, lo que nos muestra que entre estas dos variables (tamaño de muestra y longitud del intervalo) existe una correlación lineal negativa, es decir, cuando aumenta el tamaño de muestra, disminuye la longitud de los intervalos estimados.

1.2. Punto b.

Para este punto se simulo una población exponencial con parámetro $\lambda = \frac{1}{5}$ y de allí, se extrajeron las respectivas muestras de los distintos tamaños.

PARA LA MEDIA:

Utilizamos la siguiente tabla para ver mas fácilmente la comparación por tamaños de muestra, de la proporción de intervalos que atrapan la verdadera media y de la longitud promedio de los intervalos para la población exponencial.

	n=10	n=30	n=50	n=100
Porcentaje de cubrimiento	0	0	0	0
Longitud promedio del intervalo	0.2792	0.1531	0.0869	0.0893

Representación gráfica del porcentaje de cubrimiento esperado de los intervalos para la media de una población exponencial:

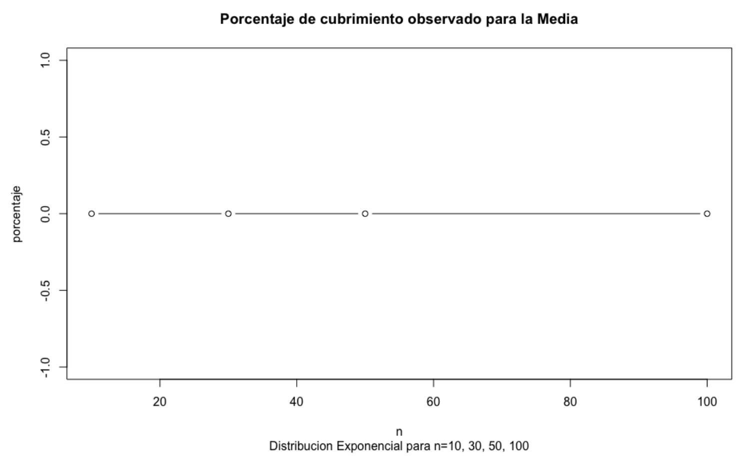

Figura 6: Gráfica del porcentaje de cubrimiento de los intervalos para μ por cada tamaño de muestra, para la población exponencial

En esta imagen podemos ver que todos los porcentajes de cubrimiento son del 0 %, al contrario que en la población normal, en la cual todos los intervalos están cerca del 95 %. Esto se debe a que la estimación por intervalos que estamos aplicando, solo se utiliza cuando la población es normal. Es decir, estas formulas no sirven para estimar μ de una población no normal. Como se ve en la imagen, nos arroja intervalos que no tienen ningún sentido con la población que estamos trabajando (en este caso, la exponencial) y que no atrapan el verdadero valor de μ.

Representación gráfica de la longitud de los intervalos para la media de una población exponencial:

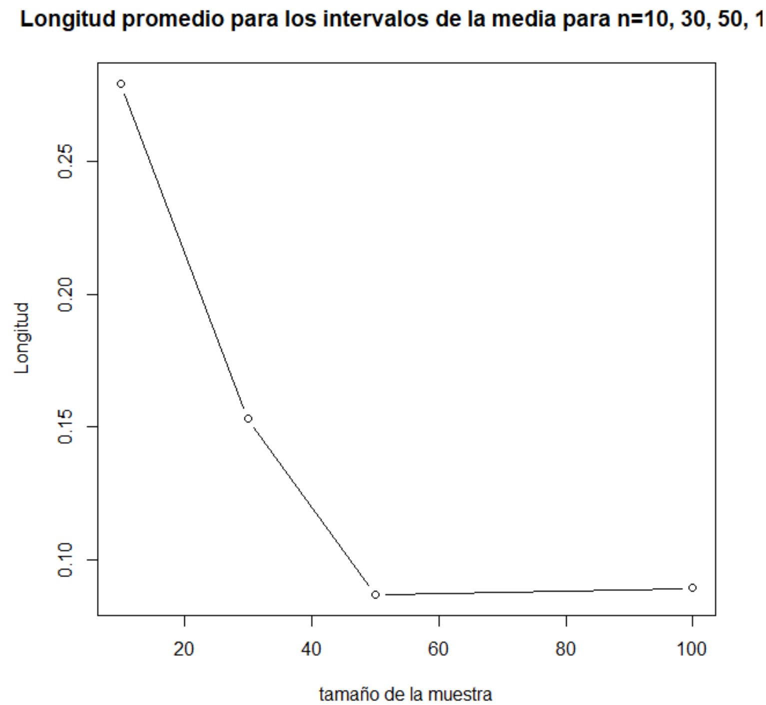

Figura 7: Gráfica de la longitud de los intervalos para μ por cada tamaño de muestra, para la población exponencial

En esta imagen podemos ver que las longitudes de los intervalos, son muy pequeñas (la máxima es de aproximadamente 0.25), lo cual nos indica que esos intervalos probablemente no tengan una alta confianza, ya que la probabilidad de que no encierren la verdadera media es muy alta. Ademas vemos que a medida que aumentan los tamaños de muestra, disminuye aun mas la longitud del intervalo, tendiendo casi a 0, volviéndose así, un estimador prácticamente puntual, el cual por obvias razones no va a contener el verdadero valor del parámetro.

PARA LA VARIANZA:

Utilizamos la siguiente tabla para ver mas fácilmente la comparación por tamaños de muestra, de la proporción de intervalos que atrapan la verdadera varianza y de la longitud promedio de los intervalos para la población exponencial.

	n=10	n=30	n=50	n=100
Porcentaje de cubrimiento	0	0	0	0
Longitud promedio del intervalo	0.1089	0.0492	0.0200	0.0293

Representación gráfica del porcentaje de cubrimiento esperado de los intervalos para la varianza de una población exponencial:

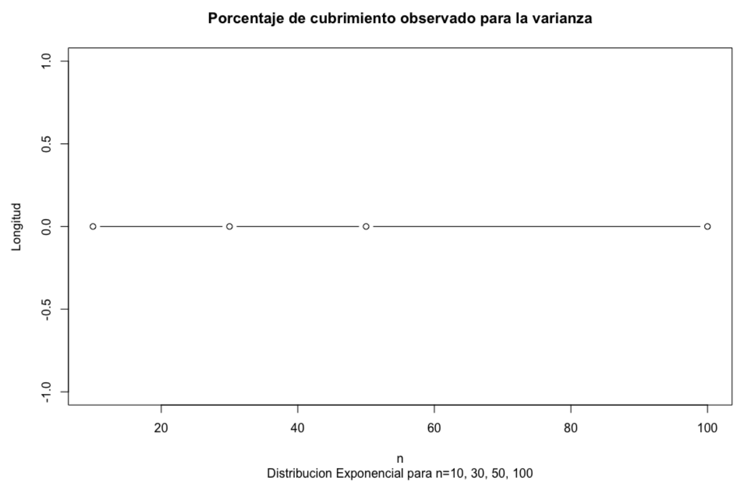

Figura 8: Gráfica del porcentaje de cubrimiento de los intervalos para σ^2 por cada tamaño de muestra, para la población exponencial

En esta imagen podemos ver que todos los porcentajes de cubrimiento son del 0 %, al contrario que en la población normal, en la cual todos los intervalos están cerca del 95 %. Esto se debe a que la estimación por intervalos que estamos aplicando, solo se utiliza cuando la población es normal. Es decir, estas formulas no sirven para estimar σ^2 de una población no normal. Como se ve en la imagen, nos arroja intervalos que no tienen ningún sentido con la población que estamos trabajando (en este caso, la exponencial) y que no atrapan el verdadero valor de σ^2.

Representación gráfica de la longitud de los intervalos para la varianza de una población exponencial:

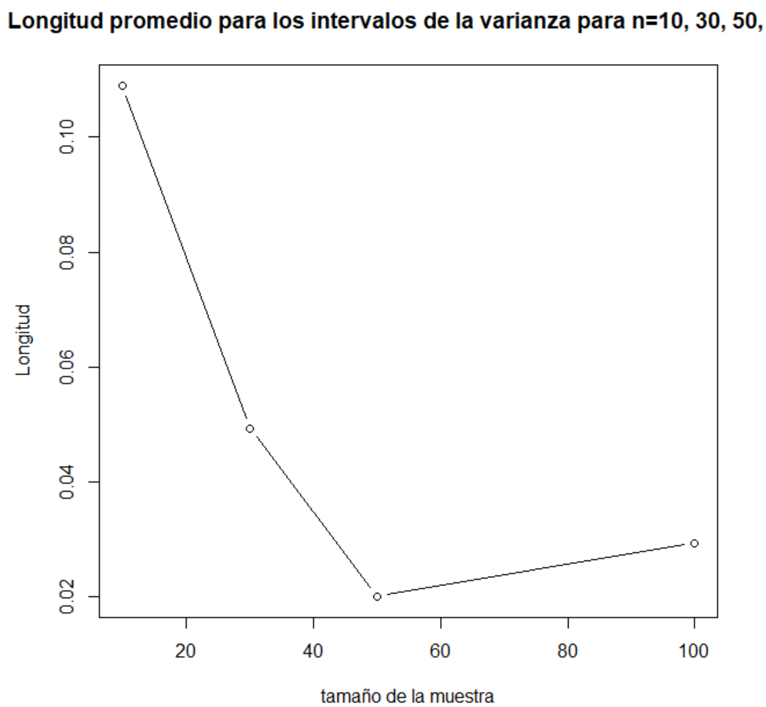

Figura 9: Gráfica de la longitud de los intervalos para σ^2 por cada tamaño de muestra, para la población exponencial

En esta gráfica, al igual que en la de para la media, vemos que las longitudes son también muy pequeñas. y Al igual que en todas las gráficas, la longitud de los intervalos tiene una correlación lineal negativo con los tamaños de muestra. Mientras mas grande sea el tamaño de muestra la longitud tiene a cero (en este caso).

CONCLUSIÓN GENERAL:
Comparando los resultados para la población normal y la exponencial, podemos ver que los intervalos son eficientes y funcionales solo para los datos que provienen de una población normal, ya que nos da porcentajes de cubrimiento cerca del nivel de confianza asignado. En cambio, los intervalos calculados para los datos que provienen de una población exponencial tienen un porcentaje de cubrimiento de 0 para cada uno de los tamaños de muestra, y para los dos parámetros (μ y σ^2). Entonces concluimos que las formulas que tenemos para este tipo de estimación, solo funcionan para datos que son provenientes de una población normal, ya que si lo aplicamos a datos muestreados de poblaciones diferentes, nos arroja intervalos que no tienen ningún sentido, y que no nos sirven en realidad como estimación de los parámetros.

2. Situación 2

Para darnos cuenta que en realidad la mayoría de personas aprueba el proyecto de fluoración del agua, debemos encontrar un intervalo de confianza de la proporción de personas que están a favor, y ver si este esta por encima del 0.5.
Entonces. Tomando los datos del enunciado tenemos:
$n = 200$, $\hat{P} = \frac{110}{200} = 0.55$ (Proporción de personas a favor), $\alpha = 0.01$ entonces $1 - \alpha = 0.99$
Ahora, para encontrar un intervalo de confianza para la proporción, aplicamos la siguiente formula:

$$IC(P)_{(1-\alpha)\,\%} = \left[\hat{P} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P}}{n}}; \hat{P} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P}}{n}}\right]$$

Reemplazando en la formula, tenemos:

$$IC(P)_{99\,\%} = \left[0.55 - Z_{0.995}\sqrt{\frac{0.55(0.45)}{200}}; 0.55 + Z_{0.995}\sqrt{\frac{0.55(0.45)}{200}}\right]$$

$$IC(P)_{99\,\%} = [0.55 - 2.5758(0.035178); 0.55 + 2.5758(0.035178)]$$

$$IC(P)_{99\,\%} = [0.4594; 0.6406]$$

INTERPRETACIÓN:

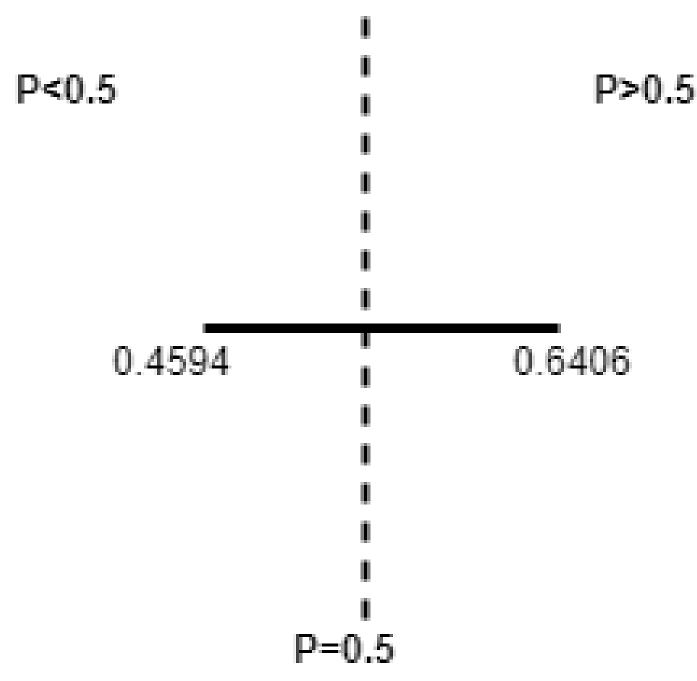

Figura 10: Interpretación intervalo para proporciones

Como asignamos una confianza del 99 % a nuestro intervalo, decimos que el 99 % de las veces que se repita el experimento, la proporción real de personas que están a favor de que se agregue fluoruro de sodio al agua va a caer en dicho intervalo (entre 0.4594 y 0.6406). Ahora, observando la gráfica, podemos ver que en el intervalo esta contenida la probabilidad de 0.5, por lo tanto, la muestra no nos da evidencia para decir que la mayoría de personas aprueba el proyecto de fluoración.

3. Situación 3

3.1. Punto a.

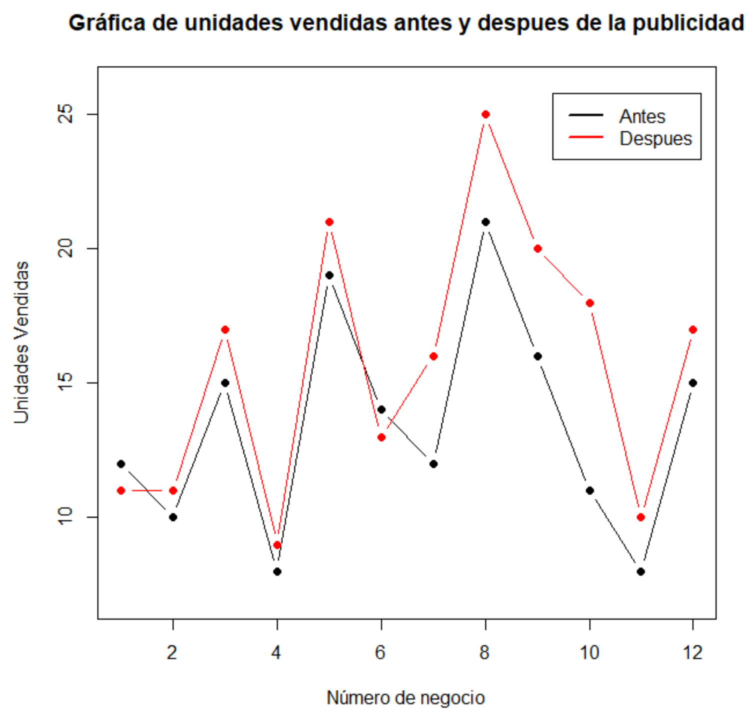

Figura 11: Gráfico de puntos y lineas por negocio de unidades vendidas, antes y después de la publicidad

En el análisis exploratorio de datos que se elaboro, encontramos que la mejor gráfica para mostrar la efectividad de la campaña publicitaria fue la gráfica de puntos y lineas. El eje x son cada una de las sucursales o negocios, y el eje y son las unidades vendidas. La linea negra representa las unidades que se vendían en dichos negocios antes de la campaña publicitaria, y la roja representa las unidades vendidas después de la campaña.
Podemos observar que en solo dos de los negocios la linea negra esta por encima de la blanca, es decir, se vendieron mas artículos antes de la campaña que después de ella; en el resto de negocios (10 negocios)la linea roja esta por encima de la negra, es decir,se vendieron mas artículos después de la campaña que antes de ella. Ademas, si miramos detalladamente las diferencias en los negocios en los cuales la linea negra esta por encima de la roja son de tan solo una unidad vendida, en cambio en los negocios en los cuales la linea roja esta por encima de la negra, hay diferencia hasta de 7 unidades vendidas.
Concluimos entonces, que con el análisis exploratorio y con la gráfica obtenida, la campaña publicitaria si es efectiva.

3.2. Punto b.

Como las muestras están relacionadas, ya que son tomadas antes y después de un tratamiento (en este caso la campaña publicitaria) a la misma población. Entonces para encontrar este intervalo debemos usar la estimación para la diferencia de medias para muestras relacionadas.
La formula que tenemos para este tipo de estimación es:

$IC(\mu_D)_{(1-\alpha)\,\%} = \bar{d} \pm t_{(1-\frac{\alpha}{2};n-1)} \cdot \frac{Sd}{\sqrt{n}}$

En este tipo de intervalo de confianza, todo se basa en la diferencia entre cada uno de los valores del antes y después del tratamiento, por tanto, para mayor comodidad encontramos las diferencias y las añadimos a la tabla:

ANTES	12	10	15	8	19	14	12	21	16	11	8	15
DESPUÉS	11	11	17	9	21	13	16	25	20	18	10	17
DIFERENCIA	1	-1	-2	-1	-2	1	-4	-4	-4	-7	-2	-2

Encontramos que: $\bar{d} = \frac{\sum_{i=1}^{12} d_i}{12} = \frac{-27}{12} = -2.25$ y $Sd = \sqrt{\frac{\sum_{i=1}^{12} (d_i - \bar{d})^2}{11}} = 2.2613$

Ahora, reemplazando en la formula, nos queda:
$IC(\mu_D)_{95\,\%} = [-2.25 \pm t_{(0.975;11)} \cdot \frac{2.2613}{\sqrt{12}}]$
$IC(\mu_D)_{95\,\%} = [-2.25 \pm 2.2009 \cdot \frac{2.2613}{\sqrt{12}}]$
$IC(\mu_D)_{95\,\%} = [-3.6880; -0.8119]$

En conclusión, un intervalo de confianza para la diferencia de medias de unidades vendidas durante un mes antes y un mes después de la campaña es (-3.6880 ; -0.8119).

3.3. Punto c.

Para interpretar el intervalo obtenido y darnos cuenta si la campaña publicitaria es efectiva o no, realizamos la siguiente imagen:

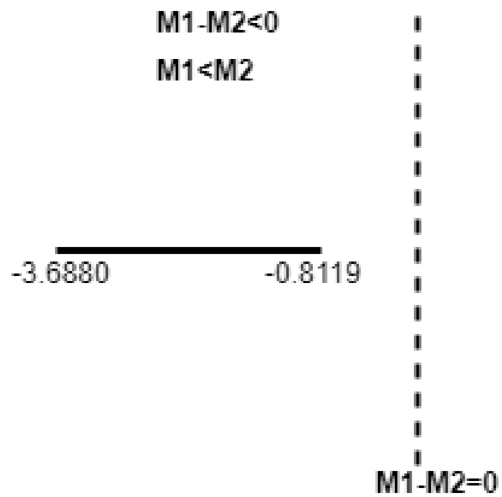

Figura 12: Interpretación intervalo para diferencia de medias en muestras relacionadas

Como asignamos una confianza del 95 % a nuestro intervalo, quiere decir que el 95 % de las veces que se repita el experimento, la diferencia real de las medias va a caer entre (-3.6880 y -0.8119). Al ver la imagen, observamos que el intervalo obtenido no contiene al cero y ademas esta debajo de el, lo que nos indica que $\mu_2 > \mu_1$, y esto nos dice que el promedio de ventas después de la campaña publicitaria es mayor que el promedio de ventas antes de la campaña.

En conclusión, podemos decir que la campaña publicitaria es efectiva ya que aumenta el promedio de ventas.

4. Situación 4

Dada una muestra de tamaño n de una población con una proporción p (desconocida) y la proporción de muestra estimada $\hat{p}$ (conocida).

4.1. Punto a.

Se construye una función para calcular el intervalo de confianza para p con un nivel de significancia α. La cual nos arroja un vector de dos elementos: el primero es el limite inferior del IC, y el segundo es el limite superior del IC.

$$IC(P)_{(1-\alpha)\,\%} = \left[\hat{P} - Z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\hat{P}(1-\hat{P}}{n}}; \hat{P} + Z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\hat{P}(1-\hat{P}}{n}}\right]$$

4.2. Punto b.

Se genera 5000 muestras de tamaño 40 de una distribución binomial con una proporción de 0.85 y un intervalo de confi anza del 95 % de la proporción de la cual arroja los siguientes resultados:

El intervalo de confianza para:
n=40, $\hat{P} = 0.85$, con un nivel de confianza del 95 % es de: $[0.8070307; 0.9929693]$
De las 5000 simulaciones , 4689 contienen la verdadera proporción de población P(0.85), teniendo entonces una cobertura del 93.78 %

De lo que se puede decir del intervalo es que al tener una longitud angosta lo hace mas preciso pues de esta forma no tendrá tanta variabilidad lo que conlleva a tener un porcentaje de cobertura bastante alto

4.3. Punto c.

El intervalo de confianza para: n=10,20,30,50,100, se podrán visualizar los resultados en la siguiente tabla, la cual contiene el tamaño de n,el intervalo de confianza, la longitud del intervalo, el contador que hace referencia al numero de veces en que de las 5000 simulaciones contienen la verdadera proporción de población P(es decir con p=0.85),por ultimo la tabla contiene el porcentaje de cobertura del intervalo de confianza.

n	Intervalo	Longitud del Intervalo	Contador	Porcentaje
n=10	[0.7140-1.0859]	0.3719	3917	78.34
n=20	[0.4991-0.9008]	0.4017	4063	81.26
n=30	[0.6153-0.9180]	0.3027	4693	93.86
n=50	[0.6416-0.8783]	0.2364	4703	94.06
n=100	[0.8163-0.9436]	0.1273	4632	92.64

Haciendo la variación en el n se puede concluir que a medida que este se va volviendo más grande el intervalo de confianza se va acotando, es decir, que este se va volviendo más angosto, lo que significa que se hace mas preciso, ya que no hay tanta variabilidad en los intervalos, dando así porcentajes de cobertura bastante altos

5. Situación 5

Como las muestras de las básculas están relacionadas, ya que una de ellas tiene sus medidas certificadas mientras que la otra se deben verificar. Entonces para encontrar un intervalo de confianza debemos usar la estimación para la diferencia de medias para muestras relacionadas.
La formula que tenemos para este tipo de estimación es:

$$IC(\mu_1-\mu_2)_{(1-\alpha)\,\%}=(\bar{x}_1-\bar{x}_2)\pm t_{(n_1+n_2-2;1-\frac{\alpha}{2})}Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

En este tipo de intervalo de confianza, se basa en la diferencia entre cada uno de los valores de las bascula en la cual se quiere verificar su estado de calibración con respecto a la báscula que se encuentra certificada, se tienen los siguientes valores:

BÁSCULA 1	11.23	14.36	8.33	10.5	23.42	9.15	13.47	6.47	12.4	19.38
BÁSCULA 2	11.27	14.41	8.35	10.52	23.41	9.17	13.52	6.46	12.45	19.35

Encontramos que:

$\bar{x}_1=12.871$ y $\bar{x}_2=12.891$

$S^2x_1=26.687$ y $S^2x_2=26.593$

$Sx_1=5.166$ y $Sx_2=5.156$

$1-\alpha=0.98$, $\alpha=0.02$

Ahora, reemplazando en la formula para hallar el intervalo de confianza, nos queda:
$IC(\mu_1-\mu_2)_{(1-\alpha)\,\%}=(12.871-12.891)\pm t_{(18;0.99)}(5.16)\sqrt{\frac{1}{10}+\frac{1}{10}}$

$IC(\mu_D)_{98\,\%}=[-0.02\pm 2.552(5.16)\frac{1}{\sqrt{5}}]$
$IC(\mu_D)_{98\,\%}=[-5.91;5.87]$

Con un nivel de confianza del 98 % se puede decir que la bascula 1 con respecto a la bascula 2 obtiene un buen rendimiento

6. Situación 6

6.1. Punto a.

Para darnos cuenta si las varianzas se pueden considerar como iguales o diferentes, debemos encontrar un intervalo de confianza para la razón de las varianzas.
Extrayendo la información del enunciado, tenemos:
$1 - \alpha = 0.98$ entonces $\alpha = 0.02$
Ahora, la formula para obtener un intervalo de confianza para la razón de varianzas es:

$$IC(\frac{\sigma_1^2}{\sigma_2^2})_{(1-\alpha)\,\%} = \left[\frac{S_1^2}{S_2^2} \cdot F_{(\frac{\alpha}{2},n_2-1,n_1-1)}; \frac{S_1^2}{S_2^2} \cdot F_{(1-\frac{\alpha}{2},n_2-1,n_1-1)}\right]$$

Hallamos S_1^2 y S_2^2:
$S_1^2 = \frac{\sum_{i=1}^{10}(x_{1i}-\bar{x_1})^2}{9} = 76875.9550$ y $S_2^2 = \frac{\sum_{i=1}^{8}(x_{2i}-\bar{x_2})^2}{7} = 42382.41037$
Reemplazando todo en la formula del intervalo, nos queda:

$$IC(\frac{\sigma_1^2}{\sigma_2^2})_{98\,\%} = \left[\frac{76875.9550}{42382.41037} \cdot F_{(0.01,7,9)}; \frac{76875.9550}{42382.41037} \cdot F_{(0.99,7,9)}\right]$$

$$IC(\frac{\sigma_1^2}{\sigma_2^2})_{98\,\%} = [1.8138 \cdot 0.14884; 1.8138 \cdot 5.61287]$$

$$IC(\frac{\sigma_1^2}{\sigma_2^2})_{98\,\%} = [0.26997; 10.18098]$$

Para interpretar mas fácilmente el intervalo obtenido, hicimos la siguiente gráfica:

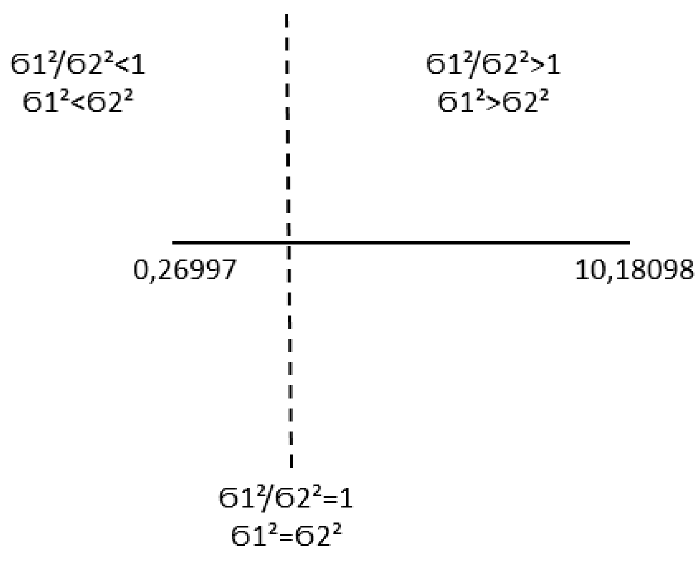

Figura 13: Interpretación intervalo para la razón de varianzas

Como asignamos una confianza del 98 % quiere decir, que el 98 % de las veces que se repita el experimento con las mismas condiciones, la razón de varianzas poblacionales va a estar entre (0.26997 y 10.18098). Si nos fijamos en la imagen, vemos que el intervalo obtenido contiene al 1.

Entonces concluimos que no se pueden considerar las varianzas poblacionales como diferentes, ya que la razón de varianzas poblacionales puede ser 1.

6.2. Punto b.

Para recomendar o no el uso del revestimiento como mecanismo complementario, encontraremos un intervalo de confianza para diferencia de medias, y así darnos cuenta si la resistencia promedio de las tuberías aumento, disminuyo, o se mantuvo igual.
Como información tenemos: $\bar{x}_1 = 2902.8$, $\bar{x}_2 = 3108.125$, $S_1^2 = 76875.9550$ y $S_2^2 = 42382.41037$

Debemos aplicar la estimación para diferencia de medias con σ_1^2 y σ_2^2 desconocidas pero iguales, ya que en el punto anterior, mostramos que con una confianza del 98 % las varianzas poblacionales pueden ser iguales.
Sustituyendo los valores en la formula, tenemos:

$$IC(\mu_1 - \mu_2)_{98\,\%} = (2902.8 - 3108.125) \pm t_{(0.99;16)} \cdot Sp\sqrt{\tfrac{1}{10} + \tfrac{1}{8}}$$

Debemos hallar Sp:

$$Sp = \sqrt{\tfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\tfrac{9\cdot 76875{,}9550+7\cdot 42382.41037}{10+8-2}} = \sqrt{61785.02922} = 248.56594$$

Ahora, reemplazando todo, nos queda:

$$IC(\mu_1 - \mu_2)_{98\,\%} = [-205.325 \pm 2.58349 \cdot 248.56594 \cdot 0.47434]$$

$$IC(\mu_1 - \mu_2)_{98\,\%} = [-509.9308; 99.2808]$$

INTERPRETACIÓN:

Para interpretar mas fácilmente el intervalo obtenido, elaboramos la siguiente gráfica:

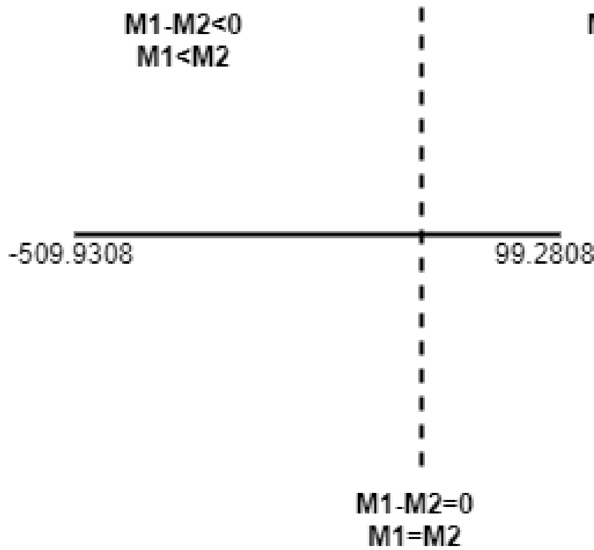

Figura 14: Interpretación intervalo para diferencia de medias en muestras independientes

Como el intervalo encontrado tiene una confianza del 98 %, decimos que el 98 % de las veces que se repita el experimento en las mismas condiciones, la diferencia de medias de resistencia reales va a caer entre -509.9308 y 99.2808. Ademas, si detallamos la imagen, podemos ver que el intervalo obtenido contiene al cero, esto quiere decir que $\mu_1 = \mu_2$ o son muy cercanos. Por lo que concluimos que el uso del revestimiento no eleva la resistencia promedio de las tuberías, y no es recomendable el uso de este como mecanismo complementario.

7. Situación 7

Para comparar los tres estimadores propuestos para el CV, se genero una población para cada distribución (normal, gamma y uniforme) y se obtuvieron 5000 muestras aleatorias para cada uno de los distintos tamaños de muestra (5,10,20,30,...,100). Las poblaciones sobre las cuales se obtuvieron las muestras fueron definidas de la siguiente manera:
1. $Normal(\mu = 200, \sigma = \sqrt{400})$
2. $Gamma(\alpha = 0.5, \beta = 100)$
3. $Uniforme(a = 200, b = 600)$

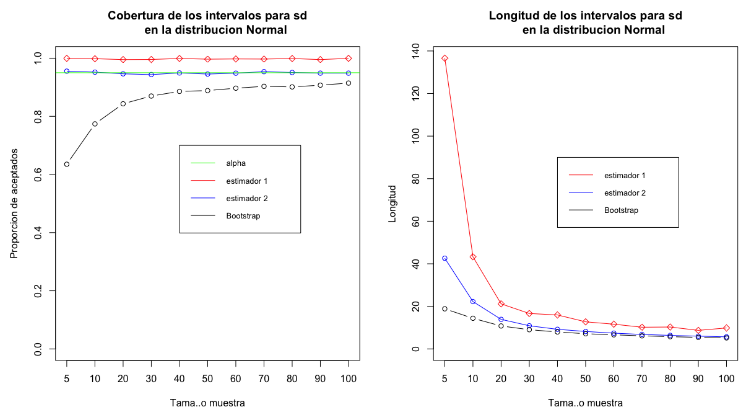

Figura 15: Proporción de cobertura y longitud promedio de los intervalos para σ de la distribución normal

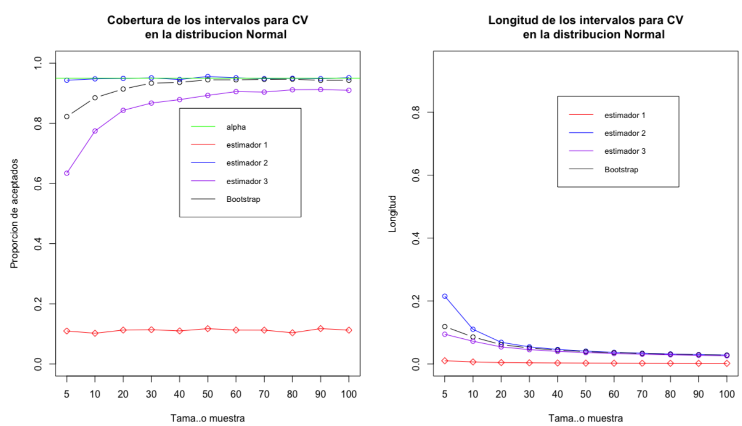

Figura 16: Proporción de cobertura y longitud promedio de los intervalos para cv de la distribución normal

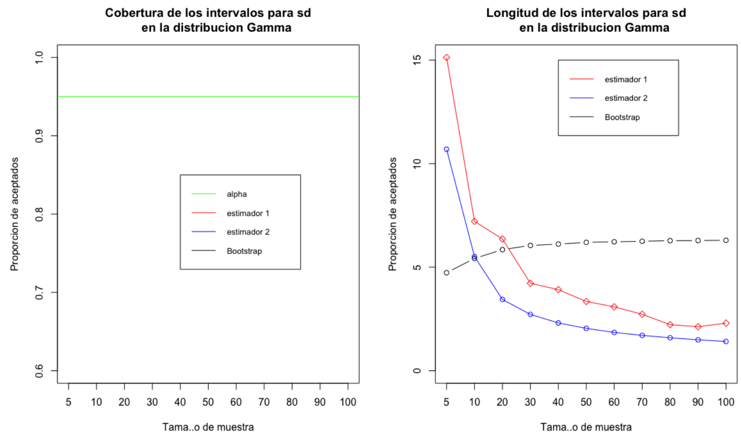

Figura 17: Proporción de cobertura y longitud promedio de los intervalos para σ de la distribución gamma

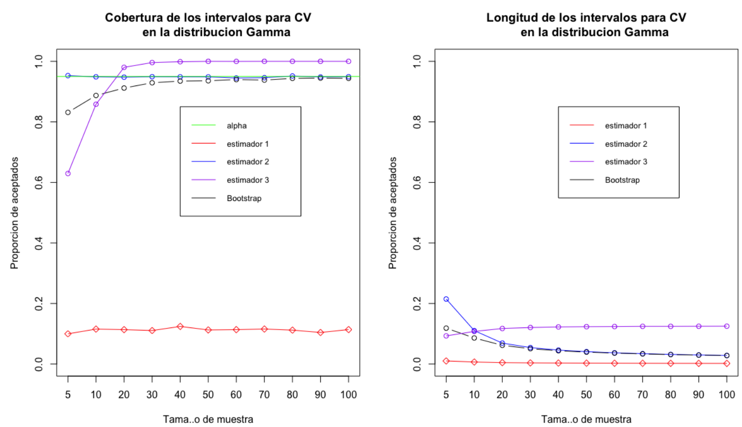

Figura 18: Proporción de cobertura y longitud promedio de los intervalos para cv de la distribución gamma

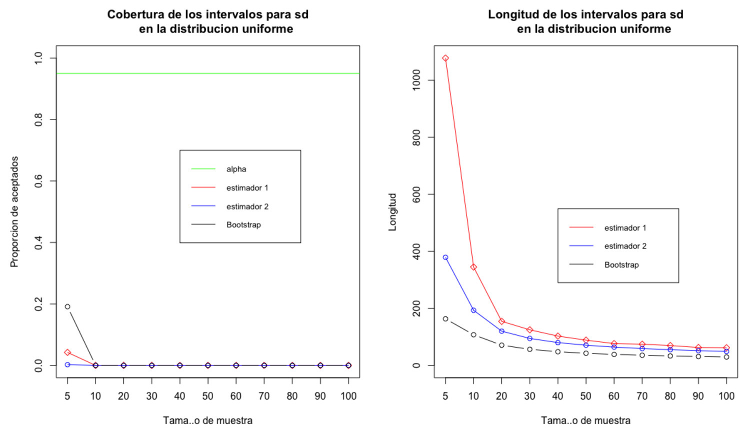

Figura 19: Proporción de cobertura y longitud promedio de los intervalos para σ de la distribución uniforme

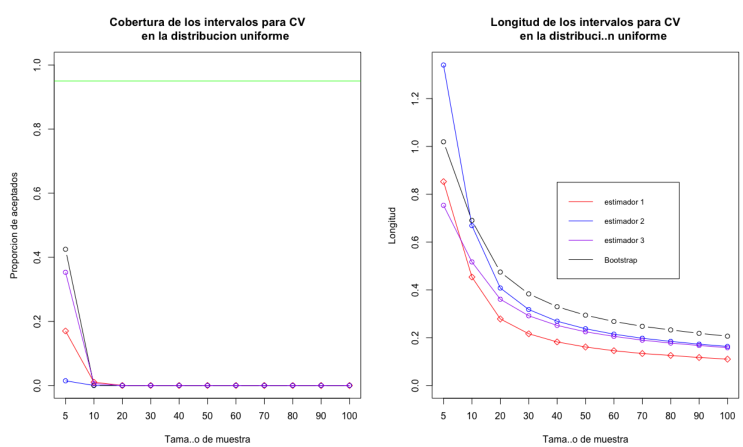

Figura 20: Proporción de cobertura y longitud promedio de los intervalos para cv de la distribución uniforme