



# MÉTODOS ESTADÍSTICOS

Comparación de medias normales asociadas al consumo de energía eléctrica en horario de invierno y verano

## Descripción breve

Universidad de Guanajuato, Departamento de matemáticas

Trabajo final de métodos estadísticos

Eloísa Díaz Francés M.

En este trabajo se modela el consumo de electricidad de la zona bajo de México con una distribución normal, esto para obtener una evidencia fuerte estadísticamente de que la media de consumo en energía en verano es mayor a la del horario en invierno.

Diego Aarón Moreno Galván

Diego.moreno@cimat.mx

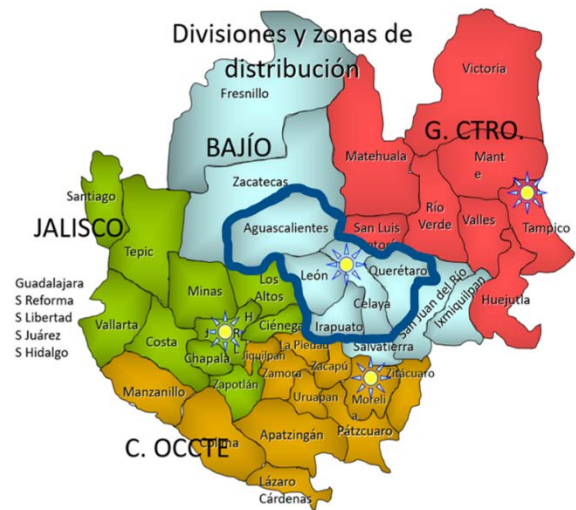
## Introducción

¿Quién no se ha preguntado si en realidad el cambio de horario de verano nos sirve para ahorrar electricidad? Esta es la motivación principal del análisis estadístico que se presenta a continuación. El análisis estadístico consiste en la comparación de las medias de muestras de consumo de energía eléctrica en la región del Bajío de México tomadas los miércoles, 10 semanas antes del cambio de horario invierno-verano y 7 semanas después de dicho cambio, bajo el supuesto de que podemos modelar el consumo como una distribución normal. Se escoge primordialmente la región del bajío porque es la zona en la que nos encontramos y se tienen condiciones climatológicas parecidas en los estados que conforman la zona. Además, las muestras de consumo se toman los miércoles de cada semana para que haya cierta independencia, puesto que los miércoles casi nunca son festivos y se gasta la misma cantidad de electricidad en teoría. La comparación del consumo se hace en base a las verosimilitudes perfiles de los parámetros  $\theta_1$  y  $\sigma_1$ , que son la media y varianza de la primera muestra (antes del cambio de horario), y de  $\theta_2$  y  $\sigma_2$ , que son la media y varianza de la segunda muestra. De esta manera nos va a interesar estimar el parámetro  $\delta = \theta_2 - \theta_1$  bajo la suposición de que desconocemos las varianzas de cada muestra usando el parámetro  $\phi = \ln \rho$ , con  $\rho = \sigma_2 / \sigma_1$ .

## Descripción de los datos

Los datos se obtuvieron directamente del Centro nacional de control de energía en México (CENACE). Nosotros solo nos vamos a enfocar en la zona de Bajío que comprende las subestaciones eléctricas de Aguascalientes, León, Irapuato, Celaya y Querétaro por conveniencia de cercanía y condiciones climatológicas similares. Ver siguiente figura.

El cambio de horario invierno-verano fue el domingo 1ro. De abril de 2018. Nuestras mediciones son dadas por el consumo de energía de las 5 subestaciones que comprende la región bajío. Éstas son medidas los miércoles (para que nuestras muestras sean independientes), durante 10 semanas antes del cambio de horario y 7 semanas después del cambio. Esto nos dice que se tomaron  $n=10$  miércoles (mediciones) antes del cambio de horario y  $m=7$  miércoles después del cambio de horario; esto también para que las condiciones del clima sean lo más parecidas. Las estadísticas suficientes asociadas son  $t_1 = 689.862$  y  $t_2 = 47626.59$  para los datos del horario de invierno y  $t_3 = 518.293$  y  $t_4 = 38391.84$  para el horario de verano.



La siguiente tabla nos muestra los datos obtenidos de las mediciones de consumo energético por cada subestación. Nosotros vamos a comparar el consumo de la región de bajío, por lo que debemos de sumar los consumos de cada subestación.

CONSUMO DEL HORARIO DE INVIERNO (kW/h)						
DÍA	CELAYA	IRAPUATO	LEÓN	AGUASCALIENTES	QUERÉTARO	SUMA (MW)
24/1	12360	13938	11525	12107	17786	67.716
31/1	12323	13637	10916	11401	17066	65.343
7/2	13228	14687	11503	11839	17826	69.083
14/2	12829	14520	11850	11534	17684	68.417
21/2	13128	14512	11836	11803	17410	68.689
28/2	12604	13857	11772	11816	17738	67.787
7/3	13449	15314	11931	12616	18067	71.377
14/3	12618	14783	11843	11685	17514	68.443
21/3	12202	15494	12287	12468	18733	71.184
28/3	13357	15704	10803	13428	18531	71.823

CONSUMO DEL HORARIO DE VERANO (kW/h)						
DÍA	CELAYA	IRAPUATO	LEÓN	AGUASCALIENTES	QUERÉTARO	SUMA (MW)
4/4	14231	16157	12545	14157	19056	76.146
11/4	13691	15597	12168	13364	18797	73.617
18/4	12880	15840	12655	13779	19331	74.485
25/4	14348	15934	12766	13448	19041	75.537
2/5	13849	15299	12259	12086	17993	71.486
9/5	13858	15067	11986	12900	18620	72.431
16/5	14239	15577	12263	13215	19297	74.591

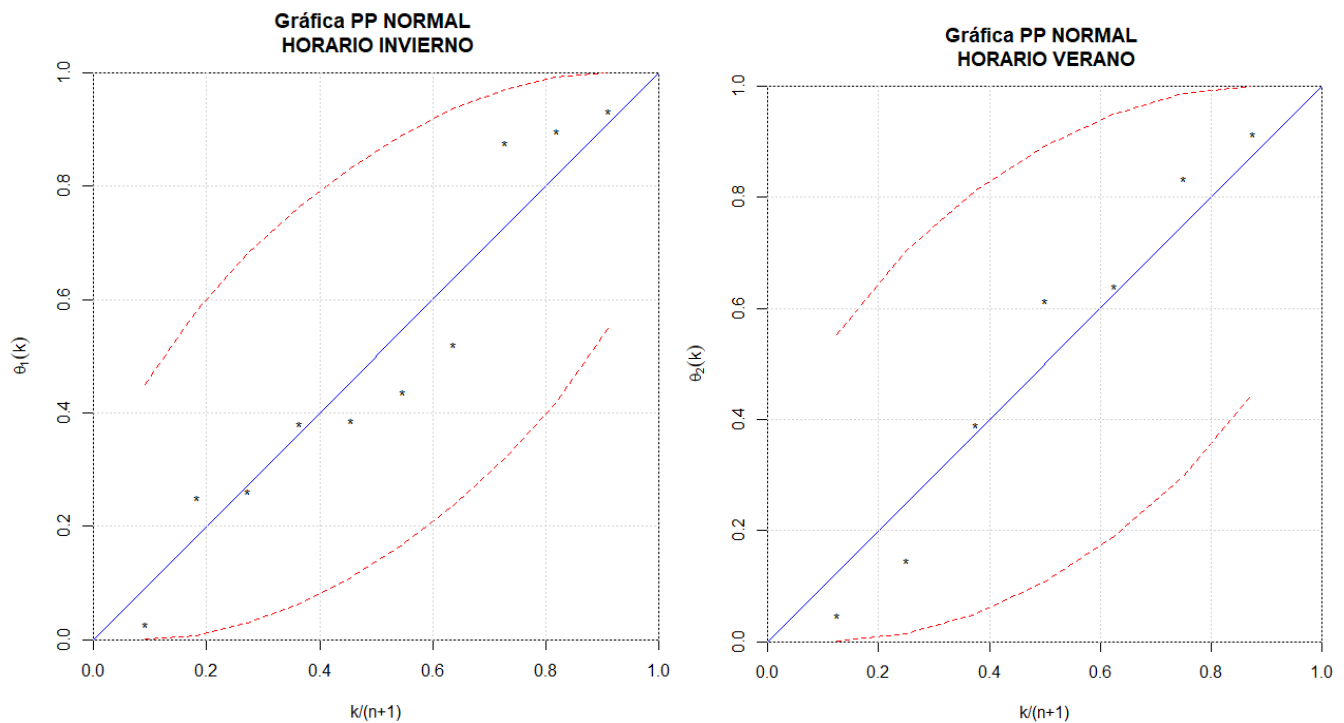
En la siguiente tabla se presentan los estimadores de máxima verosimilitud de cada parámetro y los niveles de verosimilitud exactos y asintóticos. Así mismo también se dan los intervalos de verosimilitud que cada nivel representa y la probabilidad de cobertura de cada intervalo. Como nos interesan los niveles de 95% de probabilidad, el nivel asintótico para éstos es el  $c=0.1465$ ; así mismo, para  $\rho$  nos interesa el nivel de 90% de confianza, el cual, asintóticamente es el nivel  $c=0.2585$ .

Parámetro	EMV	Nivel de verosimilitud	Intervalo de ver.	Probabilidad
$\theta_1$	68.98	0.1053	[67.56, 70.41]	0.95
		0.1465	[67.69, 70.27]	0.923
$\theta_2$	74.04	0.0887	[72.51, 75.57]	0.95
		0.1465	[72.73, 75.35]	0.919
$\sigma_1$	1.89	0.1040	[1.25, 3.34]	0.95
		0.1465	[1.28, 3.17]	0.928
$\sigma_2$	1.53	0.0891	[0.93, 3.23]	0.95
		0.1465	[0.97, 2.92]	0.917

$\delta$	2.47	0.1167	[3.14, 6.97]	0.95
		0.1465	[3.26, 6.85]	0.936
$\rho$	0.66	0.1895	[0.186, 2.583]	0.90
		0.2585	[0.211, 2.232]	0.862

### Validación del modelo normal para los datos

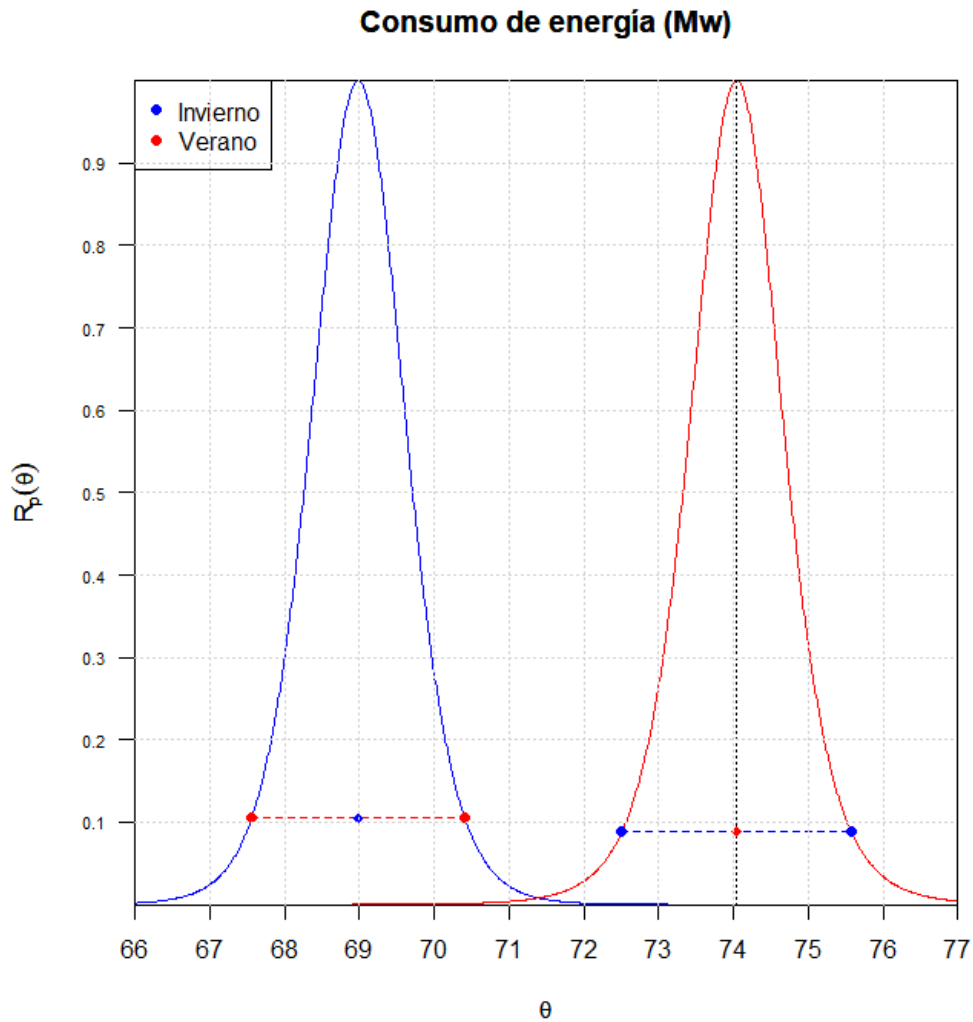
Las siguientes gráficas de probabilidad (PP) para los datos del horario de invierno y el horario de verano nos muestran que la distribución normal es un modelo razonable para éstos, pues todas las observaciones se acomodan aproximadamente de forma lineal y además ninguna observación queda fuera de las bandas de confianza de estas gráficas.



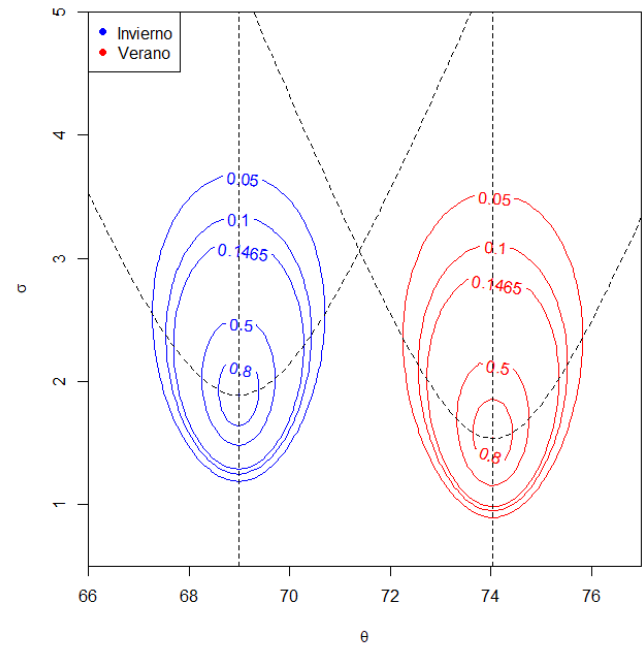
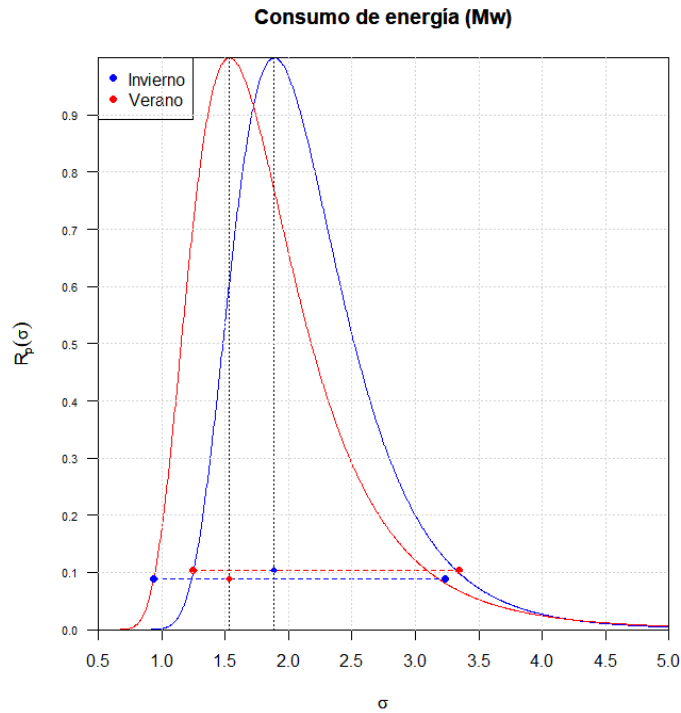
### Inferencia sobre nuestros parámetros de interés

Este ejemplo en particular es muy bonito, puesto que podemos observar que la gráfica de las verosimilitudes perfiles relativas de la primera muestra y la segunda muestra, al nivel de confianza de 95%, no existe traslape alguno de las gráficas, lo cual nos quiere decir a primera que no hay posibilidad de que haya una igualdad de medias. Claramente esto es dado por un primer análisis y a primera instancia no podríamos concluir nada sobre si son medias distintas o no, pero esto nos da

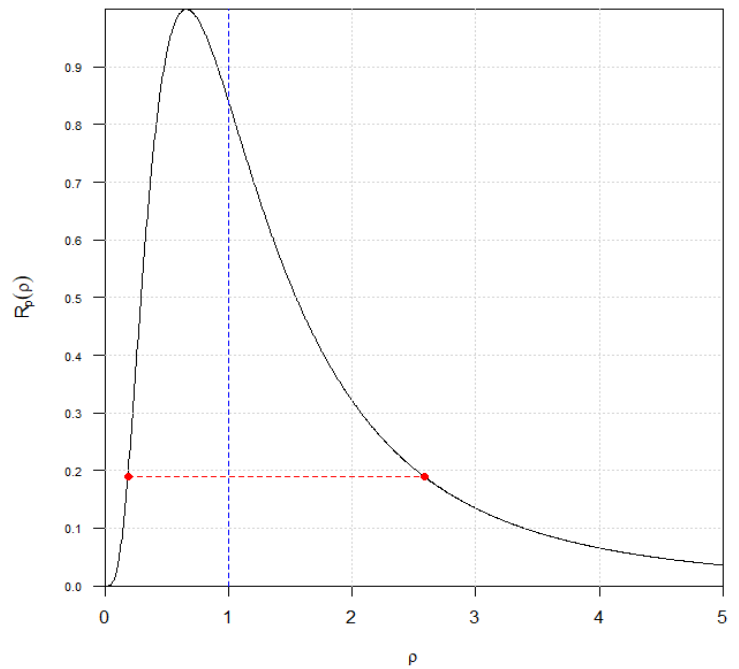
una evidencia a favor de que las medias son distintas. Observando la gráfica nos podemos dar cuenta de que la media del consumo en verano es ligeramente mayor que la de invierno. La siguiente figura nos muestra dicha gráfica.



Además, las gráficas de las verosimilitudes perfiles de la varianza ( $\sigma_1$  y  $\sigma_2$ ) de la muestra muestran un fuerte solape al nivel de confianza de 95%, lo cual nos ayuda a comparar las muestras de manera que sea válido nuestro análisis. En las siguientes figuras se muestra las verosimilitudes perfiles de las varianzas (a la izquierda) y los contornos de las medias y varianzas de las muestras (a la derecha). De los contornos podemos observar que tampoco hay un traslape desde el nivel 0.05, el cual es muy bajo. Esto es debido a que las verosimilitudes perfiles de las medias no tienen una intersección en los intervalos de confianza como antes hemos mencionado.



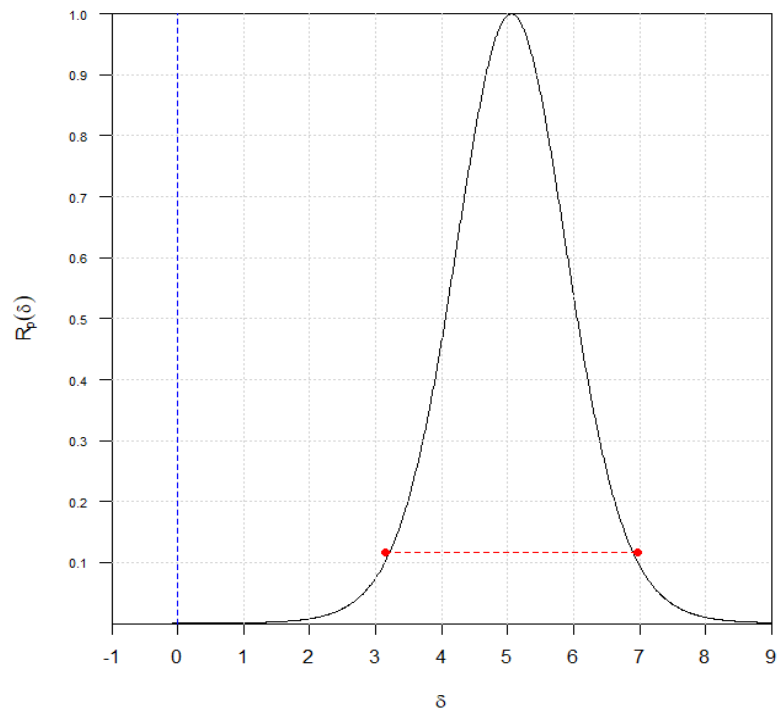
De manera similar, al graficar la verosimilitud perfil relativa del parámetro  $\rho = \sigma_2/\sigma_1$  podemos observar que el valor de 1 está sustentado en el intervalo de probabilidad 0.9 por los datos, pues tiene una verosimilitud relativa alta de  $R_p(\rho = 1) = 0.845$ . Lo cual nos dice que es muy probable que se dé una igualdad de varianzas para nuestros datos. Además de esto, también están sustentados valores más pequeños de  $\rho$ , lo cual nos dice que la varianza de la primera muestra (consumo de invierno) podría ser mayor que la segunda. Sin embargo, también están sustentados los valores de la varianza para los cuales, la varianza del consumo en verano podría ser hasta 2.5 veces más grande que en invierno. Lo que es claramente observado en las gráficas de las verosimilitudes perfiles de  $\sigma$ . La figura de la derecha nos muestra dicha gráfica.



Habíamos comentado que las verosimilitudes perfiles relativas de las medias nos mostraban que la igualdad de medias era un caso muy poco probable porque no se daba ni un pequeño solape en las gráficas. Sin embargo, vamos a profundizar el análisis.

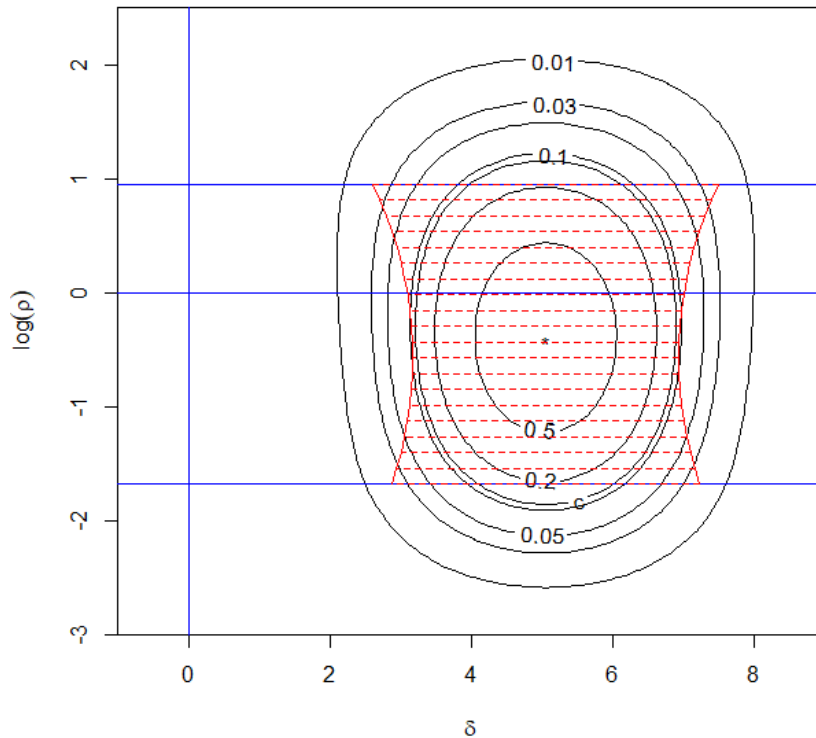
La estadística  $T_{SW}$  de Satterthwaite y Welch para probar nuestra hipótesis de que hay igualdad de medias ( $\delta = 0$ ), pues  $\delta = \theta_2 - \theta_1$ , es de 5.69, la cual es muy alta, y si las medias son iguales entonces seguiría una distribución t de Student con 14 grados de libertad. De esta manera, la probabilidad de ver algo más extremo es el p-valor y éste es 0.000487, el cual es un valor demasiado chico. Por lo tanto, existe una evidencia fuerte en contra de que las medias sean iguales, pues ésta es una probabilidad extremadamente baja. Sin embargo, nos falta corroborar lo anterior con la verosimilitud perfil de  $\delta$  y graficar los contornos de la verosimilitud perfil de  $\delta$  y  $\phi = \ln(\rho)$ .

Al momento de graficar la verosimilitud perfil de la diferencia de medias, lo que obtenemos es la gráfica de la derecha. La cual nos da una pesada evidencia de que las medias son distintas puesto que el intervalo de 95% de confianza, excluye totalmente al cero. Dicho intervalo es el que se encuentra marcado. Además, esto nos sustenta de que la media en el horario de verano es mayor que la media en el horario de invierno y la verdadera magnitud de la diferencia de medias con probabilidad de 0.95 está en el intervalo [3.14, 6.97], como ya habíamos mencionado en las tablas para el parámetro de  $\delta$ .



### Robustez de las inferencias sobre $\delta$

En la siguiente figura se muestran los contornos de verosimilitud perfil relativa de  $\delta$  y  $\phi$ . Como nuestro supuesto es que no conocemos las varianzas de nuestras muestras, tomaremos 20 valores equiespaciados dentro del intervalo de confianza del 90% para  $\phi$ . Para cada uno de estos valores de  $\phi$  graficaremos los intervalos del 95% de confianza para la diferencia de medias  $\delta$  y los marcamos con líneas rojas.



Observamos que todos los intervalos excluyen a la recta  $\delta = 0$  y el contorno correspondiente al nivel  $c$  de 95% de confianza, no interseca a la recta 0. Por lo tanto, este ejemplo sí es un caso robusto. Así que tanto la verosimilitud perfil de  $\delta$  como los contornos, nos dan una evidencia firme para decir que las medias normales de nuestros datos no son la misma. Además, los contornos hasta de nivel de verosimilitud de 0.01, también excluyen al cero lo cual nos reafirma que la probabilidad de que las medias sean iguales es bajísima.

## Conclusiones

Una conclusión ya antes mencionada, es que recabamos una prueba suficiente de la importancia de hacer el análisis estadístico con la sugerencia del criterio de Sprott y farewell. Si solamente comparamos las verosimilitudes perfiles relativas de las medias de dos conjuntos de datos puede que nos falte información para concluir si hay diferencia entre ellas o no. Sin embargo, con el análisis de robustez en nuestro modelo, logramos concluir que si existe una discrepancia en las medias con probabilidad 0.95, es decir, con certeza del 95% el consumo de energía eléctrica en la zona bajío es mayor cuando se cambia el horario de invierno a horario de verano.

Puesto que la diferencia verdadera del 95% de confianza entre el consumo de electricidad en horario de verano y de invierno está entre 3.14 y 6.97 Mw, el consumo en verano es claramente mayor que en el horario de invierno. Esto debido a que el cambio de horario se hace para aprovechar más la luz del sol, no para ahorrar electricidad, pues la gente, al salir de sus trabajos o actividades por la tarde, y todavía ver luz de sol, los incita a realizar más actividades, por ejemplo: ir al súper, al bar, o divertirse. Mientras que, si oscurece temprano, para muchas personas es sinónimo de irse a su casa, no salir, encerrarse e irse a dormir. Por lo que muchos negocios no gastan tanta energía eléctrica. También, muchas empresas aprovechan que todavía hay luz de sol para extender jornadas de trabajo, algo ilegal, pero que consume energía.

Así, en conclusión, con ayuda del análisis estadístico, el cambio de horario invierno-verano en la zona del bajío, NO sirve para ahorrar electricidad; empero sirve para que te rinda más el día, para que realices más actividades y para cuestiones comerciales.