

# Probabilidad y Estadística para Bachillerato, CIMAT, ene-jun 2019

---

## Examen Final

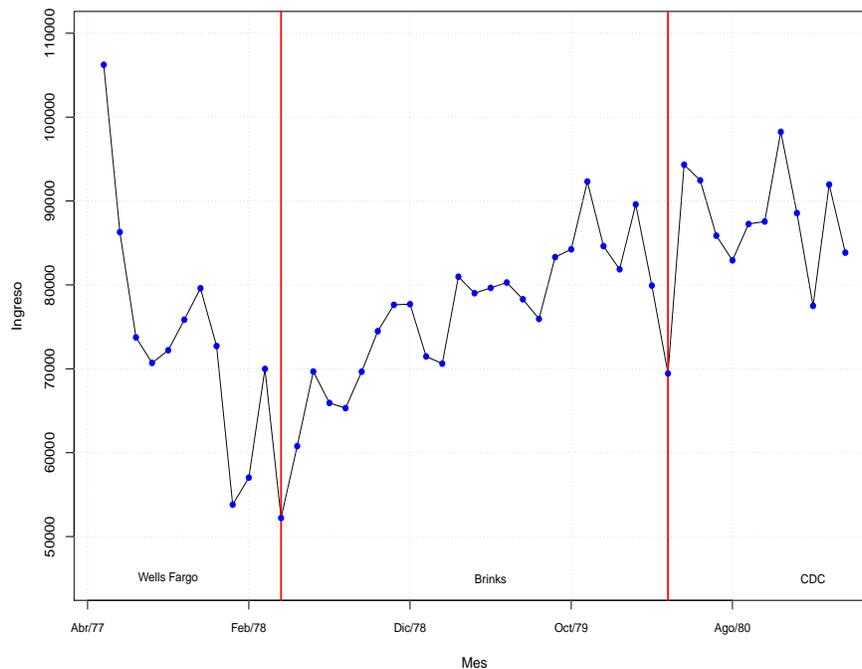
Nombre: \_\_\_\_\_

1. En Estados Unidos (y en muchos otros lados también) se tienen leyes anti-discriminatorias. En un caso legal, una mujer negra demandó a la tienda donde trabajaba por discriminación. Los antecedentes son los siguientes:

En cierta temporada navideña, una tienda departamental contrata a esta persona. Para poder ser contratada, la tienda requiere que se tome cierta prueba sobre capacidades. Sin embargo, debido a la urgencia de contratar empleados por la temporada, se le contrató de todos modos sin haber tomado antes dicha prueba. Dos semanas después, tomó la prueba y no la pasó. Consecuentemente la tienda la despidió. La mujer demandó a la tienda alegando que la prueba está sesgada pues una gran proporción de personas de raza negra la han reprobado. Los datos históricos de resultados de las pruebas son:

	Aprueba	No Aprueba	Totales
Negros	448	322	770
Blancos	240	101	341
Totales	688	423	1,111

- a) ¿Tiene razón la mujer?, ¿Cómo usaría intervalos de confianza para contestar esta pregunta?
- b) Suponga que la prueba tiene poca calidad predictiva de buena capacidad en el trabajo, esto es, si un candidato pasa la prueba no es seguro que tendrá un buen desempeño. Sin embargo, suponga que la prueba si tiene buen poder predictivo de mala capacidad, esto es, si un candidato no pasa la prueba entonces es una muy buena señal de que no tendrá buen desempeño en su trabajo (o sea, la prueba es una especie de prueba sobre habilidades mínimas).  
Bajo estas consideraciones, ¿Se afectan las conclusiones del inciso anterior?
2. Los datos contenidos en el archivo `Parquímetros.csv`, bajo la columna "Ingreso", se refieren al ingreso diario promedio por concepto de Parquímetros, durante un periodo de 47 meses en la Cd. de Nueva York. Durante cierto periodo la empresa Wells Fargo se ocupó del proceso de recolección de las monedas contenidas en cerca de 50,000 parquímetros. A partir de Mayo de 1978 el contrato fue otorgado a otra empresa llamada Brinks. Actuando en base a una denuncia anónima, se detectó que empleados de Brinks se estaban quedando con parte de lo que recaudaban. El contrato fue cancelado y se contrató a otra empresa, CDC, la cual fue supervisada de muy cerca para evitar robos.
- a) ¿Hay evidencia estadística de fraude? (Comparando el periodo de Brinks versus el periodo de CDC) (Compare intervalos de confianza para la media del ingreso diario) (Suponga Normalidad en los datos).
- b) ¿Es razonable suponer Normalidad para estos datos?. Justifique su respuesta.



```

dat = read.csv("Parquimetros.csv",header=TRUE) # 47 2
attach(dat)
n = dim(dat)[1]
plot(1:n,Ingreso,type="l",xlab="Mes",ylim=c(45000,110000),xaxt="n")
grid()
axis(1,at=c(0,10,20,30,40),labels=c("Abr/77","Feb/78","Dic/78",
                                     "Oct/79","Ago/80"),cex.axis=.8)

abline(v=c(12,36),col="red",lwd=2)
points(1:n,Ingreso,pch=16,col="blue")
text(5,45000,"Wells Fargo",cex=.8)
text(25,45000,"Brinks",cex=.8)
text(45,45000,"CDC",cex=.8)

```

3. Cierta culto religioso ha sido acusado de "comportamiento dañino para la sociedad". Se sabe que su líder pregona, entre sus adeptos, el suicidio como solución a tener dudas sobre el culto y, por alguna razón, promueve el tomar viajes largos en automóvil, con descansos mínimos. Se tienen estadísticas de 10 muertes en la comuna de 4,000 miembros, durante un periodo de 10 años por causas de suicidio o accidentes automovilísticos, mientras que en la población en general la cifra es de 13.2 por cada 100,000 habitantes por año.

Asumiendo un modelo Poisson para el número de muertes, dé un análisis de estos datos.

4. Cuando se hace una auditoría a una empresa, algunas veces los contadores, en vez de revisar cada factura, revisan a detalle solo una muestra aleatoria del total de facturas. En una ocasión, de un total de 1000 facturas, había 17 facturas fraudulentas, sin embargo ninguna de ellas apareció en la muestra de 100 que los auditores revisaron. Cuando la compañía se fue a la quiebra, un inversionista demandó a los contadores por negligencia pues basó sus decisiones de inversión en el reporte positivo de los contadores.

- a) ¿Cuál es la probabilidad de que ninguna de las 17 facturas malas apareciera en la muestra de 100?
- b) ¿Cuál sería la probabilidad correspondiente si se hubiera tomado una muestra del doble del tamaño?
- c) ¿Cuál es su conclusión? ¿Negligencia o no negligencia?
- d) ¿Que debería uno de hacer si las 1000 facturas son por cantidades muy variables?
5. Entre 1969 y 1979 se registraron 12 casos de Leucemia infantil en Woburn, Massachusetts. El número esperado, de acuerdo a la tendencia en todo el país, era de 5.3 para una región del tamaño de Woburn. Ahora bien, este pueblo esta constituido por 6 distritos o áreas geoestadísticas (AGEB's) más o menos del mismo tamaño (las AGEB's son las áreas más o menos homogéneas que se usan en los Censos para agrupar segmentos de población).
- De los 12 casos de Leucemia, se observó que 6 de ellos ocurrieron en una AGEB específica. Este hecho motivó una demanda en contra del proveedor de agua potable de esa AGEB, aduciendo que la causa del desproporcionado número de casos de Leucemia debió ser debido a contaminación del agua.
- ¿Cómo calcularía la probabilidad de que un grupo de 6 o más casos aparecieran en un mismo distrito? (bajo el supuesto de que los 12 casos ocurren puramente al azar en los distritos)
6. Considere la pregunta 11 de la hoja 2 del cuestionario de la EIC 2015. Aquí se pregunta: ¿Cuántos focos tiene esta vivienda? (datvív\$FOCOS).
- a) ¿Cuál es el número medio de focos en una vivienda en el estado de Guanajuato?
- b) ¿Cuál es la mediana?
- c) Haga un diagrama de barras frecuencia vs número de focos. ¿Cómo explicaría el comportamiento raro observado?
7. Considere la pregunta 23 de la hoja 3 del cuestionario de la EIC 2015. Aquí se pregunta: En esta vivienda, ¿Tienen auto propio (automóvil o camioneta)? (datvív\$AUTOPROP).
- a) ¿Qué proporción de viviendas encuestadas cuenta con auto propio?
- b) ¿Cuáles son los 5 municipios con la mayor proporción de viviendas con auto propio?, ¿Cuáles son los 5 con menor proporción?
8. Considere la pregunta 42 de la hoja 9 del cuestionario de la EIC 2015. Aquí se pregunta sobre el número de hijas e hijos nacidos vivos.
- a) Para mujeres en los rangos de edades: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, calcule el número promedio de hijas e hijos.
- b) Muestre gráficamente los promedios obtenidos en el inciso anterior.

---

Observaciones:

- Los problemas del 1 al 5 se refieren a cuestiones legales en las que la Probabilidad y la Estadística han tenido que decir algo acerca de la plausibilidad o no de ciertos eventos. Fueron tomados del libro:
  - Finkelstein, M.O. & Levin, B. (1990). *Statistics for Lawyers*. Springer.
- Los problemas del 6 al 8 son acerca de análisis de datos de la Encuesta Intercensal 2015 del INEGI. Los datos pueden ser obtenidos de la página del INEGI:
  - <https://www.inegi.org.mx/programas/intercensal/2015/>
  - En la pestaña de Microdatos está una lista de los estados. Descargar la opción CSV correspondiente a Guanajuato. Nos dará el archivo comprimido eic2015\_11\_csv.zip.
- Fecha de entrega: Lunes 3 de junio