

## Tarea 1 Análisis de datos categóricos

Fecha de entrega: 14 de marzo 2019

1. Lee el artículo de Tapia, José A. y Nieto, F. Javier. **Razón de posibilidades: una propuesta de traducción de la expresión odds ratio.** *Salud Pública de México.* Julio-Agosto, 1993. Vol. 35 No.4 Pág. 419-424. Elabora un resumen en media cuartilla.

2. ¿Qué es la sensibilidad (*sensitivity*) y la especificidad (*specificity*) de una prueba? ¿Qué es la curva ROC? Busca algún ejemplo para ejemplificar estos conceptos.

3. En el artículo Clopper, C. J. & Pearson, E. S. 1934. **The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial** *Biometrika*, Vol. 26, No. 4, vienen las tablas para calcular un intervalo EXACTO de confianza para p. Úsa esas tablas para calcular un intervalo al 95% y otro al 99% para una muestra tamaño  $n=50$  y  $x=9$

4. Indica el esquema de muestreo y la hipótesis nula y alternativa para cada uno de los siguientes ejemplos:

Ejemplo	Esquema de muestreo	Hipótesis nula y alternativa
a) En un hospital se decide registrar el número de parejas que lleguen en 1 año con la finalidad de someterse a un tratamiento de fertilidad. Al cabo del año, se genera una tabla en donde se registra el grupo de edad de la mujer al iniciar el tratamiento (20 a 30 y 31 a 45) y el resultado (embarazo o no embarazo).		
b) En un hospital se quiere evaluar la asociación entre edad de la mujer y resultado de un tratamiento de fertilidad. Se establece al inicio del estudio que se registraran los resultados de 722 personas. Al finalizar el tratamiento se genera una tabla en donde se registra el grupo de edad de la mujer al iniciar el tratamiento (20 a 30 y 31 a 45) y el resultado (embarazo o no embarazo).		
c) En un hospital se quiere evaluar si el resultado de un tratamiento de fertilidad cambia de acuerdo a la edad de la mujer que desea embarazarse. Para ello se deciden registrar los resultados de 350 mujeres cuya edad está entre los 20 y 30 años, y 360 mujeres cuya edad está entre los 31 y 45 años. Al finalizar el tratamiento se genera una tabla en donde se registra el grupo de edad de la mujer al iniciar el tratamiento y el resultado (embarazo o no embarazo).		

5. Los datos del ejemplo 4a) se han resumido en la siguiente tabla:

Resultado	Embarazo	No embarazo
Edad de la mujer		
20 a 30 años	26	85
31 a 45 años	146	565

5.1 Realicen “a mano” la prueba de Ji cuadrada. Concluyan.

5.2 Corrobores el resultado en R. Muestren el resultado.

5.3 Calculen la probabilidad de embarazo para mujeres de 20-30 años y la probabilidad de embarazo para mujeres de 31-45 años

5.4 Calculen el *Odds Ratio* a mano y en R. Interpreten el valor OR y los límites superior e inferior del intervalo de confianza.

5.5 Calculen el *Relative Risk* a mano y en R. Interpreten el valor RR y los límites superior e inferior del intervalo de confianza.

6. Se realizó un estudio en la India para determinar la asociación entre dos tipos de virus de papiloma humano con la aparición en las lesiones del cérvix. Realicen una prueba exacta de Fisher “a mano” y en R. Interpreten.

Cervix	Sin lesiones	Con lesiones
Tipo de Virus		
HPV 11	47	3
HPV 16	14	42

7. Comparación de métodos para generar intervalos de confianza

7.1 Generen intervalos de confianza al 99% para una p estimada de  $p=0.9$  y  $p=0.5$  con  $n=10,20,30,40,50,60,70,80, 90$  y  $100$  usando los comandos `binom.exact`, `binom.wilson` y `binom.asymp`.

7.2 Muestra los límites de los intervalos generados a través de los tres métodos

7.3 Realicen una gráfica en donde comparen los intervalos generados por los tres métodos y las dos p estimadas

7.4 ¿Cuál método elegirían? ¿Cuál método no elegirían? ¿Por qué?