

# Primer trabajo de Bioestadística para Tecnología Médica, 2015

Eliseo Martínez Herrera

27 noviembre del 2015

## Abstract

Taabajaremos con una hebra de ADN, la llamada HAR1, y buscaremos propiedades "markovianas" de longitud 1 y 2. Luego pediremos la construcción de una tabla de frecuencia para mediciones biométricas (simuladas). Y finalmente pediremos un ajuste lineal según la técnica de mínimos cuadrados a una nube de datos bidimensionales.

## 1 La hebra HAR1

Suponga usted que tenemos la siguiente hebra "pequeñísima" de ADN

*AATCGCGTAGTTA*

Y estamos interesados en conocer la frecuencia de las bases que le siguen inmediatamente a una base determinada. Para la hebra anterior si tomamos la base *A* entonces ocurre *AA* (en la posición 1), ocurre *AT* (en la posición 2) y ocurre *AG* (en la posición 9). Se realiza este procedimiento condicionado a las restantes bases, y se entregan los resultados en la siguiente tabla

	<i>A</i>	<i>T</i>	<i>C</i>	<i>G</i>
<i>A</i>	1	1	0	1
<i>T</i>	1	1	1	0
<i>C</i>	0	0	0	2
<i>G</i>	0	2	1	0

Notemos que esta tabla nos indica que aparecen 2 veces el par *GT* y una vez el par *GC*, conforme lo indica la cuarta fila. La tercera fila nos indica

que solo aparece 2 veces la combinación  $CG$  en la hebra, que usted puede comprobar (posición 4 y posición 6). Sumamos el total de los valores por fila y lo dividimos a cada entrada de la fila, obteniendo

	$A$	$T$	$C$	$G$
$A$	1/3	1/3	0	1/3
$T$	1/3	1/3	1/3	0
$C$	0	0	0	1
$G$	0	2/3	1/3	0

y así obtenemos una matriz cuyas entradas son no negativas y cuya suma por filas resultan ser 1. A esta matriz se le llama *matriz markoviana* de orden 1.

Pues bien, se pide calcular la matriz markoviana de orden 1 generada por la hebra  $HAR1$ .<sup>1</sup>

**Nota:** Para los cálculos de esta matriz, realizados en EXCEL, puede acudir a las indicaciones en el siguiente sitio en la Web

<http://www.uantof.cl/estudiomat/bioinformatica/adnexcell.wmv>

Luego, construir una tabla de frecuencia para cada base que aparece después de dos bases determinadas. Es decir, suponga que elegimos la combinación  $AA$  (de un total de 16 combinaciones posibles, ¿verdad?), entonces calculamos las frecuencias de las triadas  $AAA$ ,  $AAT$ ,  $AAC$  y  $AAG$ . Luego elegimos la combinación  $AT$  y calculamos las frecuencias de las triadas  $ATA$ ,  $ATT$ ,  $ATC$  y  $ATG$ , y así sucesivamente. Le sugerimos que construya una tabla de 16 filas (los 16 pares de bases posibles) y cuatro columnas (las cuatro bases) y en cada celda fila columna coloque la frecuencia respectiva. La ayuda para esta construcción está en el mismo sitio Web anterior.

## 2 Mediciones biométricas: bostonianos, pigmeos y dinkas

Richard Lewontin en su libro *Diversidad Humana*, Editorial Labor, primera edición:1984, trata en el capítulo 6, llamado Variación Continua, la altura

---

<sup>1</sup>Recuerde que la puede encontrar en el sitio

<http://www.uantof.cl/estudiomat/tmmt25/har1a.xls>

de los hombres adultos de Boston, pigmeos del Congo y los dinkas de África. Con las observaciones allí entregadas hemos simulado una base de datos para las alturas, en centímetros, de estos tres tipos de adultos. Esta base de datos la puede encontrar en el siguiente sitio de Internet

<http://www.uantof.cl/estudiomat/tmmt25/bospigdin.xls>

Con esta base de datos<sup>2</sup> realice lo siguiente:

- Una tabla de frecuencias para cada tipo de adultos con su respectivo gráfico
- Cálculo de medias y desviaciones estándar para cada tipo de adultos
- Cálculo de percentiles para cada tipo de adultos
- Gráfico de "caja y bigote"<sup>3</sup>, para cada tipo de adulto. (Para realizar esta caja debe calcular el primer cuartil y tercer cuartil, que son las paredes de la caja, el mínimo y el máximo, que son los terminales de los bigotes, y dentro de la caja debe marcar la mediana y el promedio)
- Conclusiones basadas en los resultados anteriores. (Ejemplo: ¿qué población es más homogénea?. etcétera)

### 3 Ajuste de un modelo lineal

Construya tres "modelos lineales" generados<sup>4</sup> por los datos "medifis" y "jóvenes dominicanos". Cada uno de los modelos lineales debe responder por lo menos las siguientes preguntas:

- ¿Cuáles son las variables predictoras (independiente) y de predicción (dependiente)? ¿qué criterio se utilizó para designarlas así?

---

<sup>2</sup>Cuidado que los datos están con el "punto" decimal. Si su EXCEL trabaja con la "coma" decimal habrá que hacer un "arreglo" de lo contrario no podrá hacer cálculos en el EXCEL

<sup>3</sup>Usted debe averiguar cómo se construye un gráfico de "caja y bigotes".

<sup>4</sup>No se considera, para medifis los datos estatura - longitud de pie. Y por lo demás tenga cuidado con el género. Recuerde que hay diferencia biométrica significativa entre hombres y mujeres, excepto en el diámetro craneal.

- ¿Es este modelo lineal un buen modelo de predicción?
- ¿Cuál es el grado de tolerancia, estimación del error, en su predicción?
- ¿Qué conclusión o conclusiones se pueden obtener de este modelo, ya sea buen modelo o mal modelo?

## 4 Indicaciones para la entrega del trabajo

- En una carpeta (con claridad el nombre de los integrantes)
- Manuscrita la **redacción de la metodología**<sup>5</sup> utilizada para cada tema.
- Gráficos y tablas pueden ser impresas y adheridas, o tambien puede anexarlas en una sección de "Anexos"
- NO TRANSCRIBA LOS DATOS ESTADÍSTICOS
- Fecha de recepción: **viernes 04 de diciembre del 2015, hasta las 12 horas** en la oficina 22 del Departamento de Matemática (oficina del profesor).

---

<sup>5</sup>Lo mínimo en la metodología utilizada es: plantear el objetivo a realizar, cómo lo van a realizar o cómo lo realizaron, las técnicas estadísticas utilizadas, el software utilizado, las conclusiones o particularidades emanadas del trabajo final, referencias de alguna bibliografía que utilizaron de apoyo y dónde utilizaron ese apoyo (fundamentalmente para la búsqueda del gráfico de "caja y bigotes")