

# TAREA 1. ESTADÍSTICA COMPUTACIONAL

23 DE NOVIEMBRE DE 2011

## 1. MEDIA TRUNCADA

La función `mean` acepta un parámetro `trim`, que toma valores entre 0 y 0.5, y que indica el porcentaje de elementos de la cola superior e inferior del vector que se descartan antes de calcular el promedio. Cuando `trim=0` se tiene el promedio habitual, mientras que `trim=0.5` es la mediana. A esto se denomina *media truncada*.

- (a)
    - Sin usar la función `mean`, programe la función `trimmed.mean` que recibe como parámetro obligatorio un vector `x` y como parámetro opcional un valor `trim` cuyo valor por omisión es 0.
    - Debe retornar la media truncada en `trim` de los valores de `x`.
    - Verifique que sus resultados coinciden con `mean` para diferentes vectores y valores de `trim`.
  - (b) Genere dos vectores de largo diez mil.
    - El vector `a` tendrá elementos distribuidos uniformemente entre -1 y 1
    - El vector `b` elementos distribuidos según una normal de media 0 y varianza 1.
    - Cargue el vector `c` del archivo `vector.c.Rdata` que se entrega por U-cursos.
  - (c) La función `sample(x, n, ...)` permite tomar una muestra de tamaño `n` de un vector `x`.
    - Genere las listas `sa`, `sb` y `sc`, de largo 200, donde cada elemento es un vector de largo 10 obtenido de un muestreo sin reemplazo de los vectores `a`, `b` y `c`.
  - (d) El comando `sapply` es útil para reemplazar un ciclo `for`.
    - Usando el comando `sapply` construya los vectores `ma`, `mb` y `mc` con las medias de cada elemento de las listas `sa`, `sb` y `sc`.
    - Dibuje un histograma para cada uno de estos vectores.
    - Sobre cada histograma dibuje un gráfico de líneas con la densidad estimada de cada vector, según el comando `density`.
- Recuerde que el histograma puede hacerse con frecuencias absolutas o relativas.
- (e)
    - Construya una lista `ta` que contenga, para cada valor en `0, 0.1, ..., 0.5`, el vector con las medias truncadas de `sa`.
    - Dibuje en un mismo gráfico, y con diferentes colores, la densidad estimada de cada vector.
    - Haga lo mismo para `sb` y `sc`.
  - (f) Presente en un único gráfico la varianza de cada elemento de `ta`, `tb` y `tc` en función del parámetro `trim` usado. ¿Qué puede comentar respecto de la varianza de `trimmed.mean` en función de `trim`?
  - (g) Generalizando lo hecho en las partes c y d vamos a verificar que la varianza del promedio de una muestra depende de su tamaño.
    - Programe una función que reciba un parámetro `n` y genere un vector con los promedios de 200 muestras de tamaño `n` tomadas del vector `b`.
    - Guarde en una lista cada vector resultante cuando `n` toma los valores 3, 6, ..., 30.
    - Grafique en una sola imagen la densidad de cada uno de esos vectores.
    - Calcule la varianza de cada vector y gráfiquela.

Hay mejor nota si en la preguntas (c) y (e) se usa `lapply` en lugar de `for`. Debe entregar un archivo R con el código usado. Incluya sus comentarios en el archivo.

## 2. MATERNIDAD

El archivo maternidad.txt contiene las fichas de los recién nacidos en una maternidad chilena.<sup>1</sup>

- (a) Cargue el archivo maternidad.txt a un data.frame. Haga un boxplot de la columna pesonac en función de sexorn.
- (b) En base al boxplot, formule una hipótesis respecto de la localización de la distribución de pesonac en función de sexorn.
  - ¿Cuál es la hipótesis nula?
  - ¿Cuál es la hipótesis alternativa?
  - Valide su hipótesis con la función t.test.
  - ¿Qué dato del resultado le indica si la hipótesis se acepta o rechaza?
- (c) Valide también su hipótesis usando los tests no paramétricos wilcox.test y kruskal.test.
- (d) Programe una función dibuja(x, y, ...) que recibe los vectores x e y, hace un gráfico de puntos, sobrepone la recta de regresión lineal entre x e y y retorna el modelo lineal resultante. Internamente conviene usar las funciones lm y abline. Si la función dibuja recibe parámetros extra, estos deben pasarse a plot.
- (e) Aplique la función dibuja a los pares (pesonac, circcraneo), (pesonac, semanagest), (pesonac, edadmadre) y (pesonac, apgar5).
- (f) Ajuste un modelo lineal que describa pesonac en función de semanagest, edadmadre y apgar5. Guarde el resultado en una variable.
- (g) Usando summary sobre el modelo lineal construido en la pregunta anterior interprete los valores de Estimate, Std. Error, t value y Pr(> |t|). ¿Hay alguna variable que se pueda descartar?
- (h) Repita el análisis anterior usando la función anova sobre el modelo lineal obtenido. Concluya.

Entregue un informe en LaTeX con los gráficos, la función dibuja y sus comentarios. Sea breve.

En ambas preguntas se valoran las respuestas claras e intuitivas por sobre las astutas o crípticas.

---

<sup>1</sup>Datos provistos por Cristian Araneda, Facultad de Agronomía, Universidad de Chile.