

# **INDESTAP: APRENDIENDO DE LOS DATOS. UN PROYECTO DE INNOVACIÓN DOCENTE EN ESTADÍSTICA APLICADA BASADO EN PROYECTOS DE INVESTIGACIÓN.**

"INDESTAP. Aprendiendo de los datos-Simulación", por Grupo de Innovación Docente en Estadística Aplicada. Departamento de Estadística e Investigación Operativa. Universitat de València, se encuentra bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 3.0 Unported](#). (2013).



---

1.- Simulación de los datos del proyecto “Análisis del efecto de la dieta en el control de la diabetes mellitus tipo 2”.

---

## 1. Simulación de los datos del proyecto “Análisis del efecto de la dieta en el control de la diabetes mellitus tipo 2”

La simulación de las variables de interés se hizo por separado para cada una de las dietas (primero se simularon los valores de las distintas variables en pacientes tratados con la dieta *LCKD* y luego en los de la dieta *LGID*) utilizando el software estadístico *R*. Cabe destacar que, en la simulación, se tuvo en cuenta la aplicación de los distintos conceptos estadísticos en los que se quería profundizar con los estudiantes. Así, con el objetivo de trabajar el concepto de intervalo de confianza, se simularon algunas variables con el modelo probabilístico normal. Otras variables se simularon con modelos probabilísticos alternativos, para utilizar contrastes de hipótesis no paramétricos.

Los detalles de la simulación de las distintas variables se muestran a continuación:

- Program. Esta variable no se simuló. Se consideraron 21 y 29 pacientes, respectivamente, como se presenta en el estudio de referencia.
- Age. La simulación de esta variable se realizó con la instrucción `rnorm(n, mean, sd)`, redondeando posteriormente el valor obtenido al entero más cercano. Los parámetros *mean* y *sd* representan, respectivamente, la media y la desviación típica. Concretando, los 21 valores de la variable *Age* de los pacientes tratados con la dieta *LCKD* se obtuvieron con la instrucción `rnorm(21, mean=51.2, sd=6.1)` y los 29 asociados a la dieta *LGID* con `rnorm(29, mean=50.0, sd=8.4)`. Los valores tomados de *mean* y *sd* son los que se presentan en el trabajo de E.C. Westman et al (2008).
- Gender. Para la simulación de esta variable se utilizó la instrucción `rbinom(n, size=1, prob=p)`, que permite simular de la distribución Bernoulli (el valor de 1 se corresponde a mujer). Para la dieta *LCKD* se tomó  $n = 21$  y  $p = 0.67$  y para la dieta *LGID*  $n = 29$  y  $p = 0.79$ .
- BMI. La simulación de esta variable se realizó de la siguiente manera:
  - o *LCKD*
    - *BMI\_w0*:  
`rnorm(21, mean=37.8, sd=6.7)`
    - *BMI\_w24*:  
`runif(21, min, max)`donde los valores de *max* y *min* se consiguieron a partir del sistema de ecuaciones:  $(min+max)/2=33.9$ ;  $(max-min)^2/12=5.8^2$ . Notar que los términos de la izquierda de ambas ecuaciones son la media y la varianza de una distribución Uniforme con parámetros *min* y *max*. Los valores de la derecha son la media y la varianza de la muestra de pacientes tratados con esta dieta que presentan los autores del trabajo.
  - o *LGID*
    - *BMI\_w0*:  
`rnorm(29, mean=37.9, sd=6.0)`
    - *BMI\_w24*:  
`rnorm(29, mean=35.2, sd=6.1)`

- Hemoglobin. Para cada una de las dietas, la variable *Hemoglobin\_w0* se simuló bajo el supuesto de normalidad mediante la instrucción de R `rnorm(n, mean, sd)`, donde *mean* y *sd* son los valores de la media y desviación típica que presentan E.C. Westman et al (2008). En concreto, para la dieta *LCKD*, *mean* = 8.8 y *sd* = 1.8; para la dieta *LGID*, *mean* = 8.3 y *sd* = 1.9.

Con el ánimo de poder profundizar en el estudio de la relación lineal entre dos variables cuantitativas, se simuló la variable *Hemoglobin\_w24* a partir de su distribución condicionada por la variable *Hemoglobin\_w0*, bajo el supuesto de que las dos variables siguen una distribución normal bivalente. Por lo tanto, una vez simulado el valor de la variable  $X_1 = \text{Hemoglobin\_w0}$  para cada uno de los individuos, se simuló la variable  $X_2 = \text{Hemoglobin\_w24}$  a partir de la expresión  $E(X_2 | X_1 = x_1) = \mu_2 + \rho\sigma_2((x_1 - \mu_1)/\sigma_1)$ , con  $\rho = 0.87$ . Para los pacientes asociados a la dieta *LCKD*,  $\mu_2 = 7.3$  y  $\sigma_2 = 1.5$ ; para los de la dieta *LGID*,  $\mu_2 = 7.8$  y  $\sigma_2 = 2.1$ .

- FastGluko. Las variables *FastGluko* antes y después de la dieta, es decir *FastGluko\_w0* y *FastGluko\_w24*, también se simularon bajo el supuesto de normalidad. Las instrucciones de R que se utilizaron fueron:
  - *LCKD*
    - *FastGluko\_w0*: `rnorm(n=21, mean=178.1, sd=72.9)`
    - *FastGluko\_w24*: `rnorm(n=21, mean=158.2, sd=50.0)`
  - *LGID*
    - *FastGluko\_w0*: `rnorm(n=29, mean=166.8, sd=63.7)`
    - *FastGluko\_w24*: `rnorm(n=29, mean=150.8, sd=47.4)`

Como en los casos anteriores, los valores de *mean* y *sd* son los mismos que presentan los autores del trabajo citado.

- FastInsulin. Las variables *FastInsulin*, antes y después de la dieta, es decir, *FastInsulin\_w0* y *FastInsulin\_w24*, se simularon con las siguientes instrucciones:
  - *LCKD*
    - *FastInsulin\_w0*: `runif(n=21, min, max)`
    - *FastInsulin\_w24*: `runif(n=21, min, max)`
  - *LGID*
    - *FastInsulin\_w0*: `runif(n=29, min, max)`
    - *FastInsulin\_w24*: `rnorm(n=29, mean=12.6, sd=6.5)`

Los valores max y min se calculan de la misma forma con la que se ha trabajado para el caso de la variable BMI.

- HDLCholesterol. En este caso la simulación se ha hecho de la siguiente manera:
  - *LCKD*
    - *HDLCholesterol\_w0*: `rnorm(n=21, mean=44.0, sd=8.7)`
    - *HDLCholesterol\_w24*: `runif(n=21, min, max)`

Los valores max y min se calculan de la misma forma con la que se ha trabajado para el caso de la variable BMI.

- *LGID*
  - *HDLCholesterol\_w0*:  
`rnorm(n=29, mean=48.7, sd=11.8)`
  - *HDLCholesterol\_w24*:  
`rnorm(n=29, mean=48.7, sd=10.1)`
- *VLDLCholesterol*. Para esta variable, se utilizó:
  - *LCKD*
    - *VLDLCholesterol\_w0*:  
`rnorm(n=21, mean=37.3, sd=14.9)`
    - *VLDLCholesterol\_w24*:  
`runif(n=21, min, max)`

donde los valores de *max* y *min* se consiguieron a partir del sistema de ecuaciones:  $(min+max)/2=27.3$ ;  $(max-min)^2/12=15.2^2$ .
  - *LGID*
    - *VLDLCholesterol\_w0*:  
`rnorm(n=29, mean=27.7, sd=13.2)`
    - *VLDLCholesterol\_w24*:  
`runif(n=29, min, max)`

donde los valores de *max* y *min* se consiguieron a partir del sistema de ecuaciones:  $(min+max)/2=24.4$  ;  $(max-min)^2/12=12.3^2$ .
- *Triglycerides*. Esta variable también se simuló, tanto antes como después de cada una de las dietas, con el supuesto de normalidad. En concreto, el código de *R* utilizado fue:
  - *LCKD*
    - *Triglycerides\_w0*:  
`rnorm(n=21, mean=210.4, sd=10.3)`
    - *Triglycerides\_w24*:  
`rnorm(n=21, mean=142.9, sd=76.9)`
  - *LGID*
    - *Triglycerides\_w0*:  
`rnorm(n=29, mean=167.1, sd=125.7)`
    - *Triglycerides\_w24*:  
`rnorm(n=29, mean=147.8, sd=128.5)`

Como en los casos anteriores, los valores *mean* y *sd* son los que se presentan en el trabajo citado.
- *Change\_med*. Esta variable fue simulada a partir de la distribución Bernoulli mediante la instrucción de *R* `rbinom(n, size=1, prob=p)`. Siguiendo el trabajo de E.C. Westman *et al* (2008), para la dieta *LCKD* se tomó  $n = 21$  y  $p = 0.95$ ; para la dieta *LGID* asumimos  $n = 29$  y  $p = 0.62$ . Los valores iguales a 1 fueron asignados a aquellos pacientes que experimentaron un mayor descenso en el nivel de hemoglobina A1c.
- *Adverse\_effect*. La simulación de esta variable también se realizó a partir de la distribución Bernoulli. En concreto, se utilizó la instrucción `rbinom(n, size=1, prob=p)`, siendo  $p = 0.53$  en el caso de la dieta *LCKD* y  $p = 0.46$  en la dieta *LGID*.