

## PRACTICA 9: inferencia sobre dos poblaciones. Enunciados.

### Objetivos

- Hacer inferencia (estimación por intervalos de confianza y contraste de hipótesis) sobre dos poblaciones para la media, la varianza y la proporción.
- Comprobar las condiciones que permiten hacer inferencia sobre los parámetros antes mencionados.
- Interpretar los resultados obtenidos. En particular,
  - Combinar la información que proporcionan los intervalos y el p-valor asociados al contraste de una hipótesis estadística.
  - Distinguir entre diferencia estadísticamente significativa y científicamente significativa.

El fichero Pima.tr se refiere a mujeres de la etnia Pima de al menos 21 años de edad que viven cerca de Phoenix, Arizona. Se realizó a cada una de ellas un test de diabetes de acuerdo con los criterios de la Organización Mundial de la Salud (WHO); los datos fueron recogidos por el Instituto Nacional para la Diabetes y Enfermedades Digestivas y de Riñón de Estados Unidos. Las variables medidas fueron

1. npreg: número de embarazos.
2. glu: concentración de glucosa en sangre tras un test oral de tolerancia a la glucosa (mg/dm).
3. bp: presión diastólica (mm Hg).
4. skin: grosor del pliege de la piel en el triceps (mm).
5. bmi: índice de masa corporal, en  $\text{kg/m}^2$ .
6. ped: diabetes pedigree function. Función que asigna la probabilidad de padecer diabetes a partir de la historia familiar.
7. age: edad, en años.
8. type: Yes or No, para diabetes, de acuerdo con los criterios de la WHO.

## Enunciados

Empieza por leer el fichero Practica07-pima.csv y guardar su contenido en la variable pima. Puedes usar la función `read.table()` de R (no olvides fijar correctamente la carpeta de trabajo)

```
pima = read.table(file = "Practica07-pima-tr-BS.csv", sep = ";", header = TRUE, dec = ".")
```

o bien usar el botón `Import dataset` de RStudio.

Visualiza las primeras filas de la tabla

```
head(pima, 4)

  npreg glu bp skin  bmi  ped age type
1     5  86 68  28 30.2 0.364 24  No
2     7 195 70  33 25.1 0.163 55 Yes
3     5  77 82  41 35.8 0.156 35  No
4     0 165 76  43 47.9 0.259 26  No
```

- Se quiere comparar si la dispersión en la presión arterial es la misma entre las indias a la que se les diagnosticó diabetes y las que no:
  - ¿Se puede afirmar, con un nivel de confianza del 95%, que la dispersión (varianza) es la misma?
  - ¿Qué relación (numérica) hay entre las varianzas?
- Determina si se puede considerar que el valor medio del grosor pliegue en el triceps (al nivel de significación del 10%) es mayor entre los individuos a los que se les diagnosticó diabetes y a los que no.
- Supón (por suponer) que el índice de masa corporal de cada individuo se ha reducido un 7% tras cinco años. Se trata de una disminución significativa? Trabaja al 5% de significación. Pista, se trata de comparar `pima$bmi` con `pima$bmi * 0.93`.
- Un estudioso afirma que el nivel medio de glucosa de los primeros 20 individuos de la tabla es menor que el de los 20 últimos. Contrasta dicha afirmación.
- Se quiere analizar la (posible) asociación entre las variables tener diabetes diagnosticada (`type`) y el índice de masa corporal (`bmi`). Para ello, analiza la diferencia de proporciones de indias con diabetes entre los individuos por debajo del primer cuartil y por encima del tercer cuartil respecto de la variable `bmi`. En concreto, considera los siguientes individuos.

```
# por debajo del primer cuartil
table(pima$type[pima$bmi <= quantile(x = pima$bmi, 0.25)])

No Yes
46  4

# por encima del tercer cuartil
table(pima$type[pima$bmi > quantile(x = pima$bmi, 0.75)])

No Yes
26  23
```

Comenta los resultados.