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 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

Para leer los datos, teclea y ejecuta el código

```
library(MASS)
pima = Pima.tr
```

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
```

1. **Se quiere comparar si la disparsión en la presión arterial es la misma entre las indias a la que se les diagnosticó diabetes y las que no**:

   (a) **¿Se puede afirmar, con un nivel de confianza del 95%, que la dispersión (varianza) es la misma?**

   (b) **¿Qué relación (numérica) hay entre las varianzas?**

2. **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**.

3. **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.**

4. **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 percentil 30 y por encima del 70 (percentiles arbitrarios) 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, 0.3)])


    No Yes
    54   7

   # por encima del tercer cuartil
   table(pima$type[pima$bmi > quantile(x = pima$bmi, 0.7)])


    No Yes
    33  27
   ```

   Comenta los resultados.

5. **Vamos a trabajar ahora con el peso de varios ratones antes y después de recibir cierto tratamiento. Instala el paquete `datarium`.**

   ```
   install.packages("datarium")
   ```

   activa el paquete

```r
library(datarium)

## Warning:  package 'datarium' was built under R version 4.3.3
```

Guarda los datos en la variable `datos`

```r
datos = mice2
```

Contrasta la hipt́esis de que el tratamiento no afecta al peso de los ratones. Trabaja al 5% de significación.