

# Práctica 2: Contrastes de hipótesis

Módulo de Modelos Lineales.  
Máster de Bioestadística, Universitat de València.

Miguel A. Martinez-Beneito

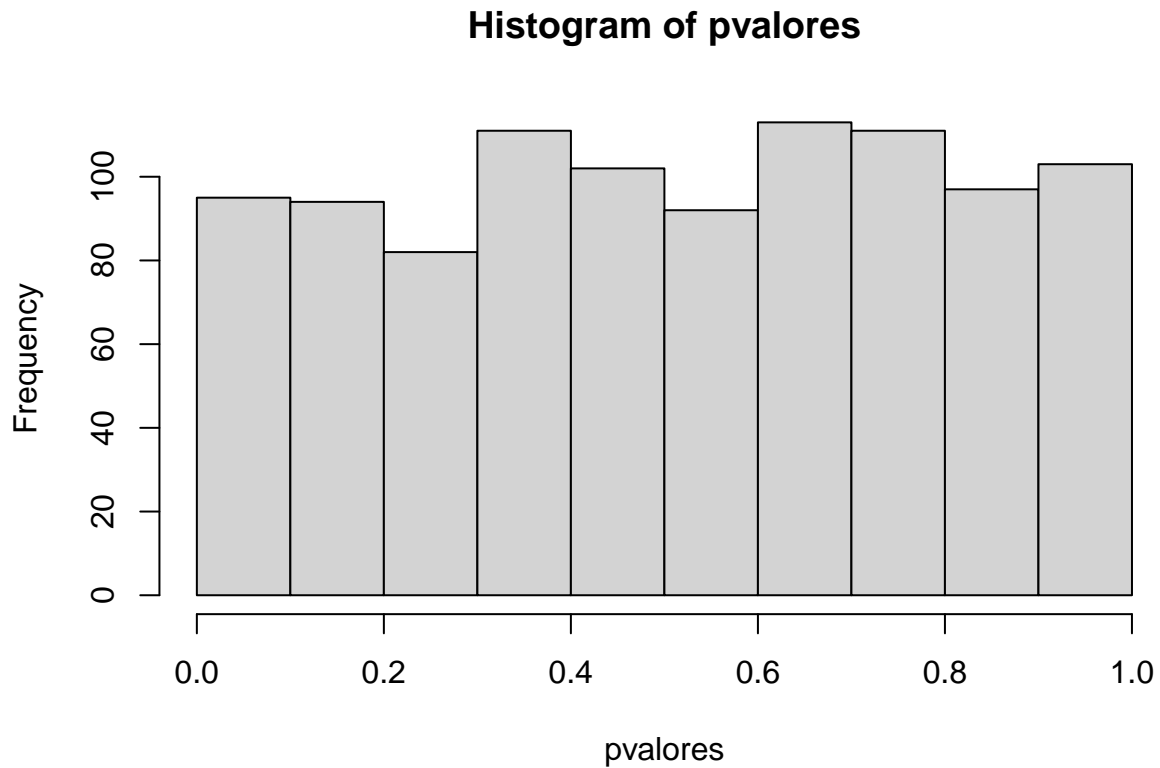
## Tareas

1. Retomamos la tarea 3 de la práctica anterior en la que valorábamos el porcentaje de veces en las que un intervalo de confianza contiene el verdadero valor del parámetro. En este caso vamos a hacer algo parecido pero con contrastes de hipótesis.
- Genera 1000 muestras de una distribución Normal standard, cada una de ellas de tamaño 100. Para cada muestra contrasta si la media poblacional de la población de la que provienen puede ser 0. Utiliza una significación  $\alpha = 0.05$  para tus conclusiones.
- Utilizando los datos generados en el apartado anterior construye un histograma de los p-valores que has obtenido en ese caso ¿Que puedes decir, a partir de lo obtenido, sobre la distribución de P-valores en contrastes de hipótesis cuando la hipótesis nula es cierta?
- Repite el primer apartado de esta tarea para distribuciones Normales de media  $0.05 \cdot (1:10)$  y desviación típica 1. Contrasta en todos esos casos si los datos pueden provenir de una distribución Normal de media 0. Por último, construye una curva con la proporción de veces que has encontrado que  $\mu \neq 0$  para cada uno de los valores que has considerado para la media de las muestras. Relaciona los resultados obtenidos con el error de tipo II para cada uno de los valores de la media de las muestras.

```
# Cálculo de P-valores
set.seed(1)
datos <- matrix(rnorm(1e+05), ncol = 100, nrow = 1000)
pvalores <- apply(datos, 1, function(x) {
  t.test(x)$p.value
})
sum(pvalores < 0.05)
```

```
## [1] 44
```

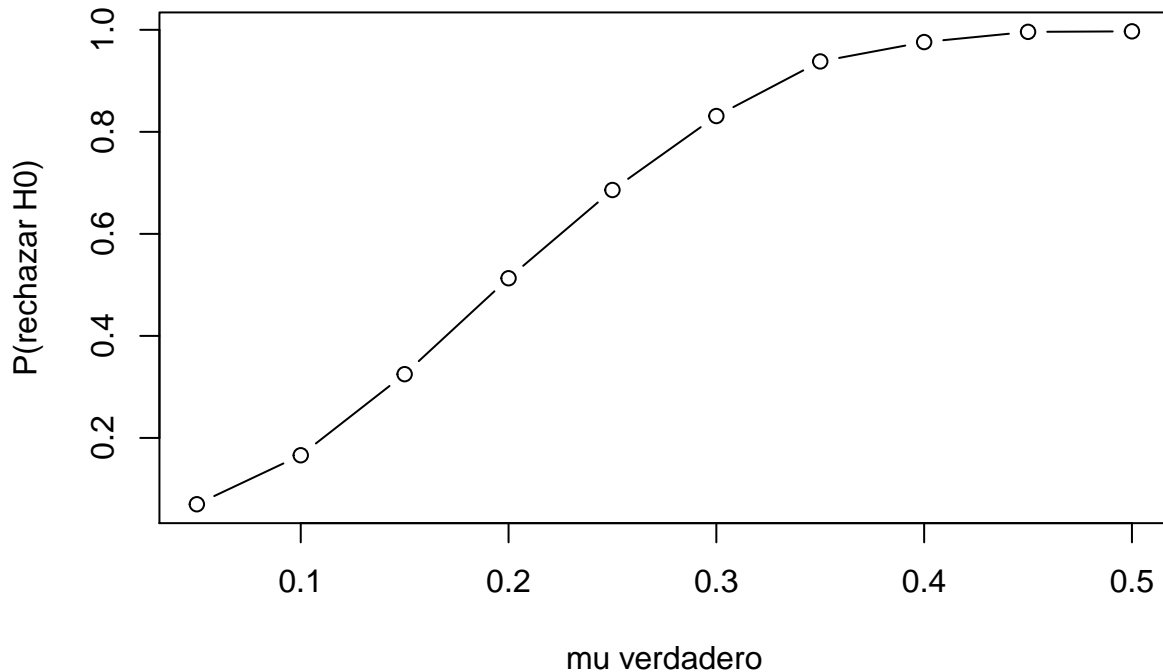
```
# Histograma P-valores
hist(pvalores)
```



```
# La distribución de los P-valores cuando H0 es cierta es uniforme

# P-valores cuando H0 false
set.seed(1)
pvalores2 <- matrix(nrow = 1000, ncol = 10)
for (i in 1:10) {
  datos <- matrix(rnorm(1e+05, mean = 0.05 * i), ncol = 100, nrow = 1000)
  pvalores2[, i] <- apply(datos, 1, function(x) {
    t.test(x)$p.value
  })
}
plot(0.05 * (1:10), apply(pvalores2, 2, function(x) {
  mean(x < 0.05)
}), xlab = "mu verdadero", ylab = "P(rechazar H0)", main = "Curva de potencia",
type = "b")
```

## Curva de potencia



*# Cada uno de los puntos anteriores coincidiría con 1-P(error de tipo  
# II) para cada valor de mu.*

2. El banco de datos **glucosa** disponible en **glucosa.Rdata** dispone de los datos de glucosa en sangre de distintos individuos con diabetes incipiente. Dichos individuos se han dividido en tres grupos, según la dieta que se le ha administrado durante el último mes: aquellos que han seguido una dieta normal, el grupo de dieta baja en carbohidratos y el grupo de dieta baja en azúcares. Vamos a querer hacer distintas comparaciones entre los grupos mencionados.

- Compara, de la forma que te parezca más oportuna, el valor medio de glucosa en sangre después de seguir la dieta (variable **GDespues**) entre el grupo bajo en carbohidratos y el grupo bajo en azúcares.
- Compara, de la forma que te parezca más oportuna, el valor medio de glucosa en sangre después de seguir la dieta entre el grupo bajo en azúcares y el grupo sin restricciones dietéticas.
- Queremos ahondar ahora en más detalle en las cuestiones anteriores. Para ello plantéate, de la manera que te parezca más adecuada, si se ha producido descenso, o no, en el nivel medio de glucosa en cada uno de los grupos de individuos que hemos observado.

*# Comparación (bilateral) entre los grupos de carbohidratos y azúcares*  
`with(Glucosa, t.test(x = GDespues[Dieta == "H. Carbono"], y = GDespues[Dieta ==  
"Azúcares"])))`

```
##
## Welch Two Sample t-test
##
## data:  GDespues[Dieta == "H. Carbono"] and GDespues[Dieta == "Azúcares"]
## t = -0.59393, df = 53.299, p-value = 0.5551
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
## -6.127295 3.327295
## sample estimates:
## mean of x mean of y
## 110.4 111.8

# Comparación (unilateral) entre los grupos de azúcares y dieta normal
with(Glucosa, t.test(x = GDespues[Dieta == "Azúcares"], y = GDespues[Dieta ==
  "Normal"], alternative = "less"))

##
## Welch Two Sample t-test
##
## data: GDespues[Dieta == "Azúcares"] and GDespues[Dieta == "Normal"]
## t = -3.5771, df = 57.686, p-value = 0.0003559
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
## -Inf -4.935991
## sample estimates:
## mean of x mean of y
## 111.8000 121.0667

# Comparaciones (unilaterales) pareadas antes-después para cada uno de
# los grupos
with(Glucosa, t.test(x = GAntes[Dieta == "Azúcares"], y = GDespues[Dieta ==
  "Azúcares"], alternative = "greater", paired = TRUE))

##
## Paired t-test
##
## data: GAntes[Dieta == "Azúcares"] and GDespues[Dieta == "Azúcares"]
## t = 16.301, df = 29, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 8.450033 Inf
## sample estimates:
## mean of the differences
## 9.433333

with(Glucosa, t.test(x = GAntes[Dieta == "H. Carbono"], y = GDespues[Dieta ==
  "H. Carbono"], alternative = "greater", paired = TRUE))

##
## Paired t-test
##
## data: GAntes[Dieta == "H. Carbono"] and GDespues[Dieta == "H. Carbono"]
## t = 20.667, df = 29, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 10.06505 Inf
## sample estimates:
## mean of the differences
## 10.96667

with(Glucosa, t.test(x = GAntes[Dieta == "Normal"], y = GDespues[Dieta ==
  "Normal"], alternative = "greater", paired = TRUE))

##
```

```
## Paired t-test
##
## data:  GAntes[Dieta == "Normal"] and GDespues[Dieta == "Normal"]
## t = -0.55636, df = 29, p-value = 0.7089
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -1.081064      Inf
## sample estimates:
## mean of the differences
##                -0.2666667
```

3. Discute la veracidad de las siguientes afirmaciones:

- Si encontramos una diferencia significativa en un contraste de hipótesis bilateral, necesariamente volveremos a encontrar esa diferencia significativa si repetimos el contraste de forma unilateral.

*La afirmación es falsa. Si el pivote cayera en la cola que favorece a la hipótesis nula el p-valor unilateral sería  $1-(p\text{-valor bilateral}/2)$ , que será un valor bastante grande.*

- Si en un contraste encontramos diferencias significativas para  $\alpha = 0.01$  también la encontraremos para  $\alpha = 0.05$ .

*Efectivamente, si encontramos diferencias a nivel 0.01 es que el p-valor es inferior a dicho valor, por tanto también será inferior a 0.05. Por tanto encontraremos también diferencias significativas para dicho valor de  $\alpha$*

- Si en un contraste encontramos diferencias significativas para  $\alpha = 0.05$  también la encontraremos para  $\alpha = 0.01$ .

*Falso. Simplemente tenemos que pensar en un p-valor de 0.03.*

- En un contraste de hipótesis, cuando rechazamos la hipótesis nula consideramos probada la hipótesis alternativa.

*Efectivamente, cuando rechazamos la hipótesis nula nos vemos abocados a considerar la hipótesis alternativa como cierta.*

- En un contraste de hipótesis, cuando no rechazamos la hipótesis nula la damos por probada.

*Falso. Cuando no rechazamos la hipótesis nula es que los datos no muestran suficiente evidencia en su contra, pero no podemos concluir que sea cierta.*