

Explorando el Transcriptoma con Datos de Expresión Genética

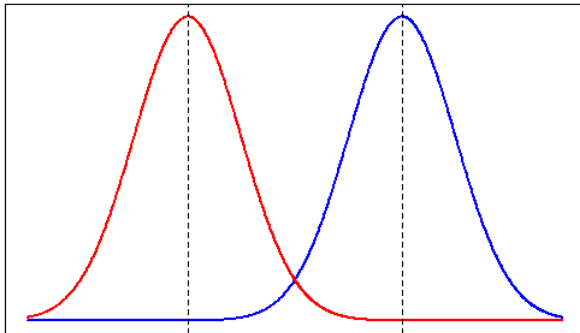
Datos de Expresión Genética

Yered Pita-Juárez

8/1/2015

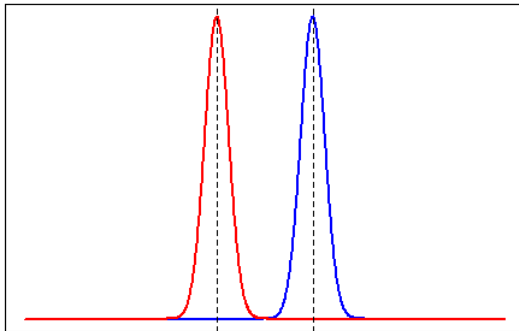
Prueba de t

- Comparar 2 grupos
- Determinar si las medias de dos grupos son significativamente diferentes



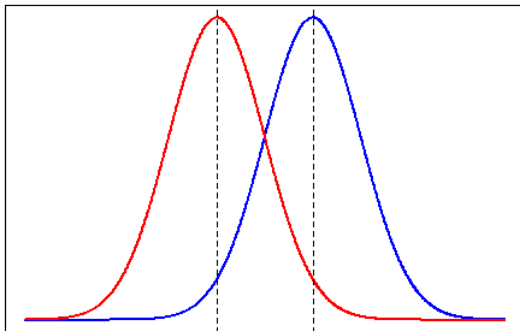
Prueba de t

- ¿Qué tan diferentes pueden ser?



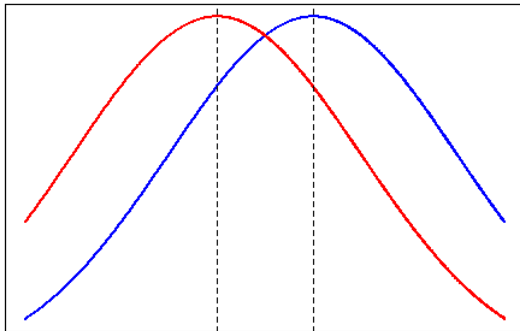
Prueba de t

- ¿Qué tan diferentes pueden ser?

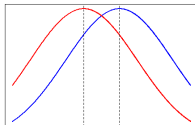
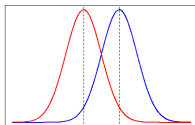
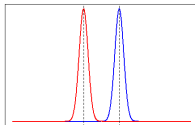


Prueba de t

- ¿Qué tan diferentes pueden ser?



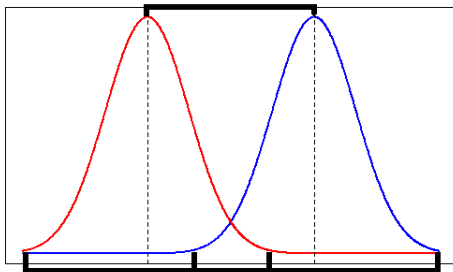
Prueba de t



Prueba de t

- La diferencia entre las medias es la misma
- Hay que tomar en cuenta la variabilidad de cada grupo
- Señal: diferencia entre las medias
- Ruido: la variabilidad de los grupos
- estadística T

$$T = \frac{\text{señal}}{\text{ruido}} = \frac{\bar{X}_1 - \bar{X}_2}{\text{sd}(\bar{X}_1 - \bar{X}_2)}$$



Prueba de t

- H_0 : las medias son iguales vs. H_1 : las medias son diferentes
- $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$
- μ_1 y μ_2 son las medias verdaderas de los grupos
- Solo tenemos observaciones para cada grupo
- Estimamos las medias usando los promedios \bar{X}_1 y \bar{X}_2

Prueba de t

- ¿Cómo decidimos entre H_0 y H_1 ?
- Valor p: si H_0 es cierta, la probabilidad de observar una estadística al menos tan extrema como la observada
- Mientras mas pequeño sea el valor p, la evidencia en contra de H_0 es mayor
- Usamos el criterio α , rechazamos H_0 si

$$p \leq \alpha$$

- Usualmente $\alpha = 0.05$
- En R, usa la función `t.test`
- `t.test(x, y)` compara las medias de los vectores `x` y `y`

Ejercicio

- Usa `rnorm` para generar dos vectores con la misma media y aplica la prueba de t usando `t.test`
- Usa `rnorm` para generar dos vectores con medias diferentes y aplica la prueba de t usando `t.test`



Ejercicio

- Descarga el archivo `CdeC.RDS` en tu `working directory`
`https://dl.dropboxusercontent.com/u/21912429/CdeC/CdeC.RDS`
- Carga los datos en R usando
`CdeC = readRDS("CdeC.RDS")`
- `CdeC` es una lista con las edades de dos clubes de ciencia
- Calcula la edad promedio para cada club usando `mean`
- Determina si la edad promedio es significativamente diferente usando `t.test`

Comparaciones Múltiples

- Usar la prueba t para comparar muchas variables entre dos grupos
- ¿Cuántas veces vamos a rechazar H_0 ?
- Si usamos un valor α de 0.05, vamos a rechazar H_0 en el 5% de los casos al azar
- Sin ajustar los valores p nuestros resultados no son confiables

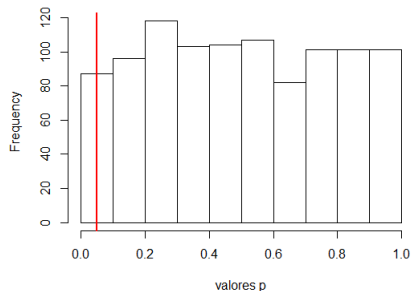
Simulación

- Vamos a generar 2 vectores con la misma media usando `rnorm`
- Vamos a usar la prueba de t para comparar las medias
- En este caso, H_0 siempre es cierta
- La desición correcta es no rechazar H_0
- Usando $\alpha = 0.05$ vamos a estar equivocados en el 5% de los casos

```
sim <- function(n=10){  
  x1 = rnorm(n)  
  x2 = rnorm(n)  
  tmp = t.test(x1,x2)  
  tmp$p.value  
}  
res = replicate(1000,sim(n=30))
```

Simulación

```
hist(res,main="",xlab="valores p")  
abline(v=0.05,col="red",lwd=2)
```



Ajuste

- Métodos para ajustar los valores p
- En este caso, al ajustarlos no deberíamos tener ningún valor p menor a 0.05

```
res_adj = p.adjust(res)
hist(res_adj,xlim=c(0,1),main="",xlab="valores p ajustados")
abline(v=0.05,col="red",lwd=2)
```

