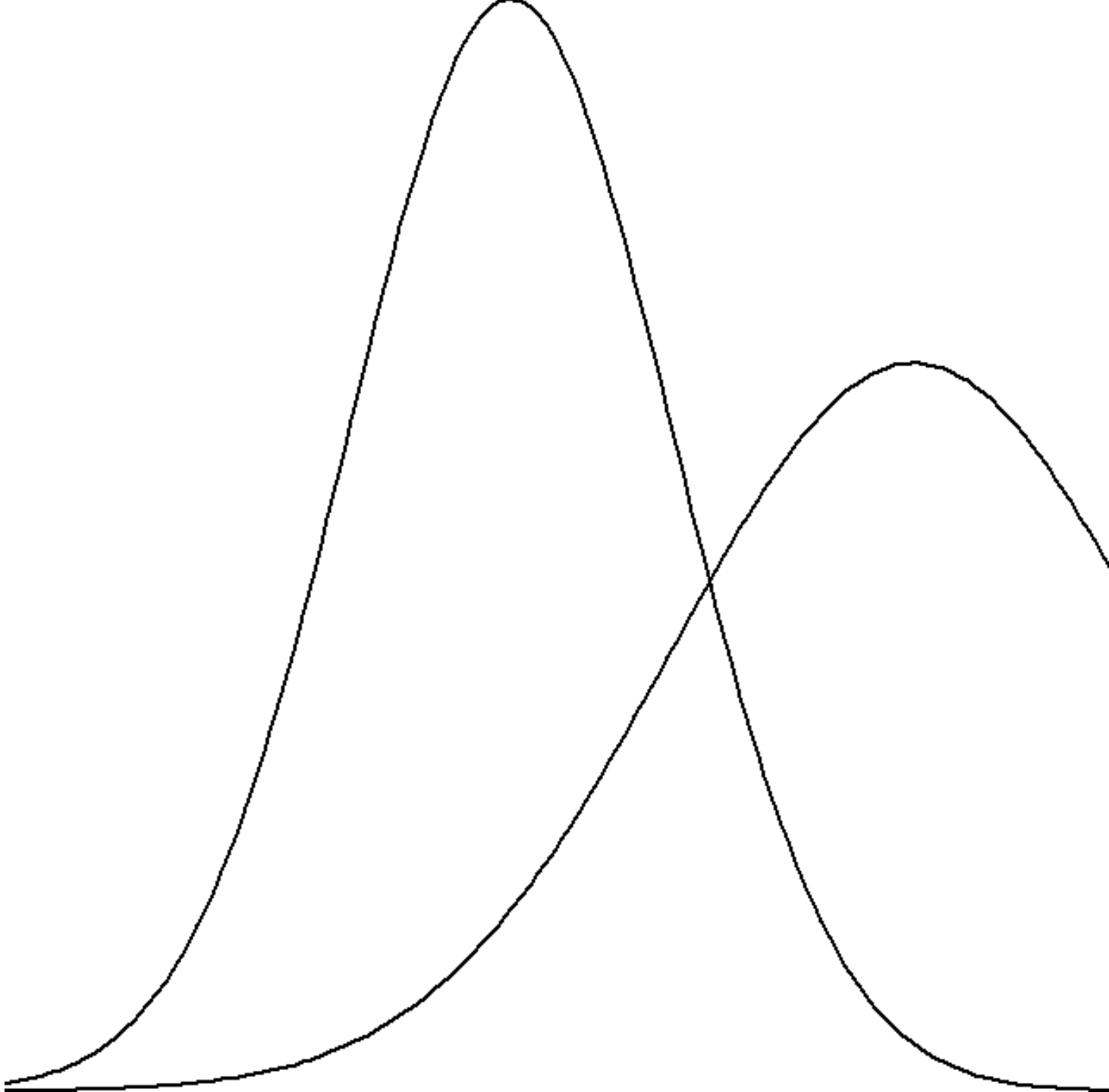
# Test de comparación de densidades (Paquete sm)

UXIO MERINO, XOEL MONTES & BORJA SOUTO



- 1. **Problema**:
  - Comparar distribuciones entre grupos va más allá de medias/varianzas (ej: formas complejas)
- 2. **Solución propuesta**:
  - Métodos no paramétricos con ***sm.density.compare*** : kernel + tests de permutación
- 3. **Aplicación real**:
  - Datos de supervivencia en tumores cerebrales (*ISLR2/BrainCancer*)

# Hipótesis y Estadístico de Contraste

- 1. Hipótesis:

$$H_0: \quad f_1(y) = f_2(y) = \cdots = f_k(y) \quad \forall y$$
$$H_1: \quad \exists\, i, j, y \mid f_i(y) \neq f_j(y)$$

- 2. Estadístico ISE:
  - Para 2 grupos:

$$T = \int \left( \hat{f}_1(y) - \hat{f}_2(y) \right)^2 dy$$

  - Para $k > 2$ grupos:

$$T = \sum_{i=1}^{k} n_i \int \left( \hat{f}_i(y) - \hat{f}_{\text{global}}(y) \right)^2 dy$$

# Implementación en *sm.density.compare*

- Selección del ancho de banda óptimo para cada grupo
- Selección de ancho de banda común: $h_{\text{común}} = \left(\prod_{i=1}^{k} h_i\right)^{1/k}$ (media geométrica)
- Estimación kernel con núcleo gaussiano: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$
- Test de permutación: $(\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I(T_b \geq T_{\text{obs}}))$
  - Redistribución aleatoria de etiquetas ⟶ p-valor empírico.

# Bandas de Referencia y Visualización

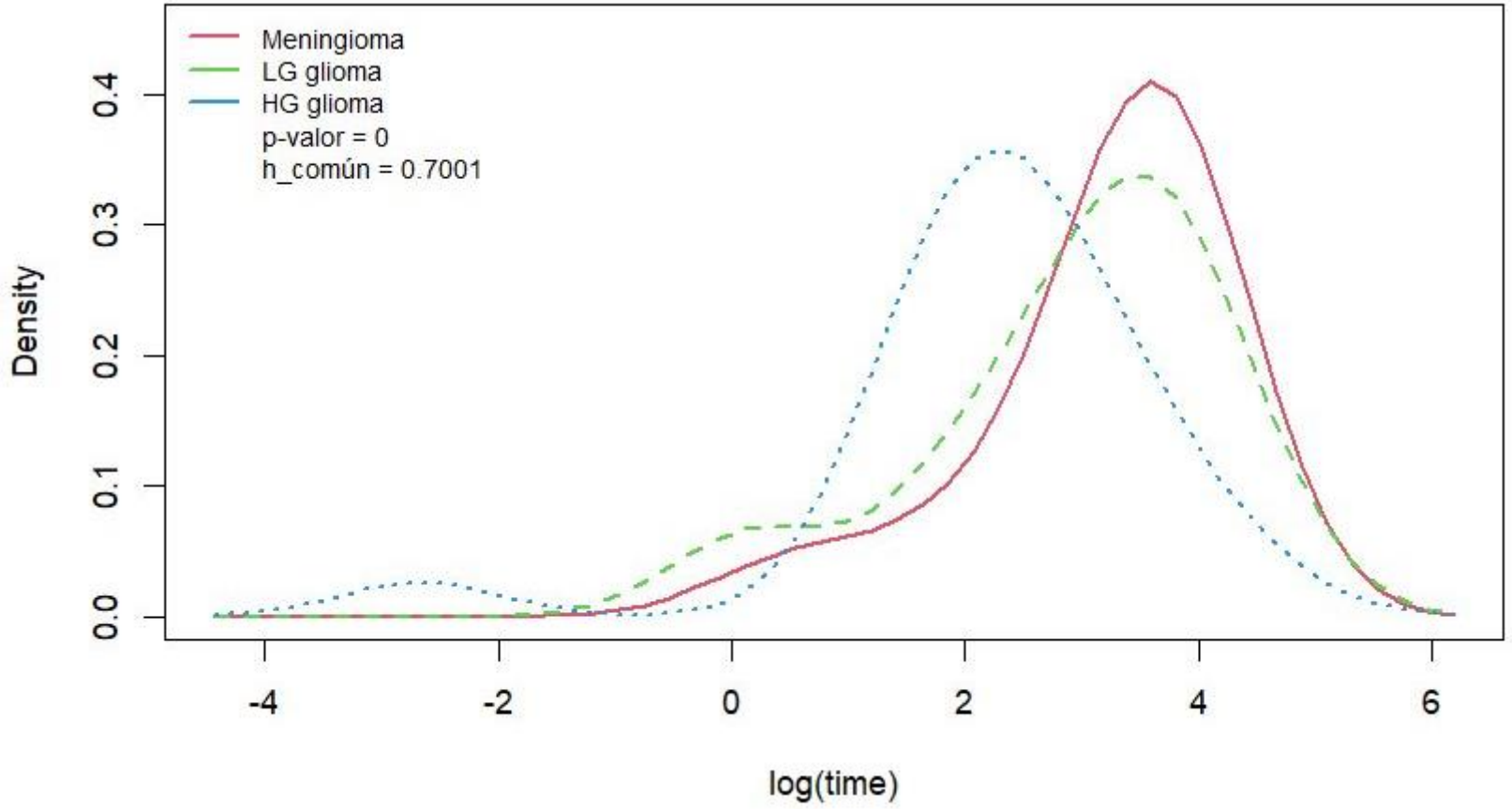
- Solamente cuando se comparan 2 grupos
- Transformación estabilizadora de la varianza $\mathrm{Var}[\sqrt{\hat{f}(y)}] \approx \frac{R(K)}{2nh_{\text{común}}}$
- $R(K)$ representa la integral del cuadrado del núcleo
- Las bandas de confianza se construyen transformando de vuelta a la escala original los intervalos derivados para las densidades transformadas

# Propiedades teóricas

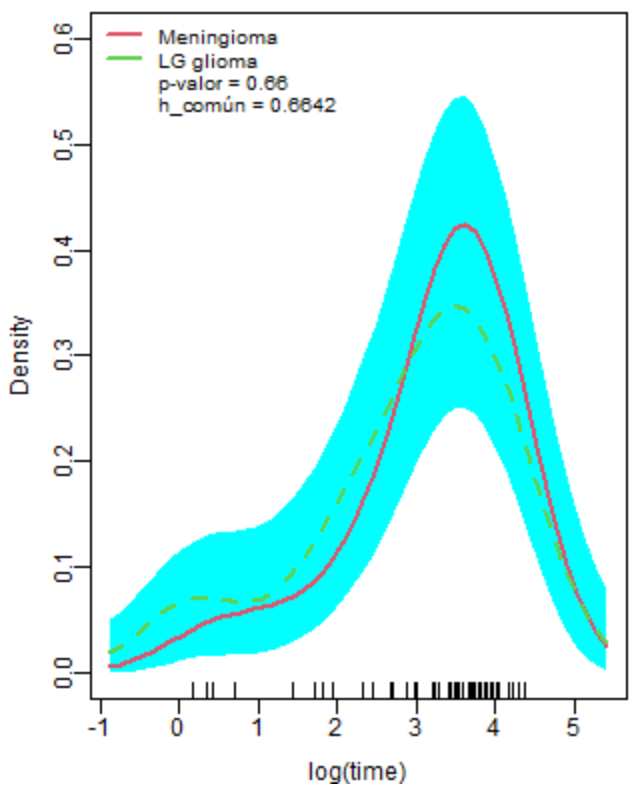
- Sesgo:
  - El uso de un h común asegura que bajo $H_0: E[\hat{f}_i(y) - \hat{f}_j(y)] = 0$
- Varianza asintótica: $\mathrm{Var}[\hat{f}(y) - \hat{g}(y)] \approx \frac{R(K)}{nh}(f(y) + g(y)), \quad \text{donde } R(K) = \int K^2(u)du$
- Consistencia: $h \to 0$ y $nh \to \infty$ cuando $n \to \infty$
- Optimalidad MISE asintótico mínimo bajo kernels simétricos no negativos (ej: gaussiano)
- Invarianza bajo transformaciones: $\hat{f}_{T(Y)}(y) - \hat{g}_{T(Y)}(y) \approx \hat{f}_Y(T^{-1}(y)) - \hat{g}_Y(T^{-1}(y))$

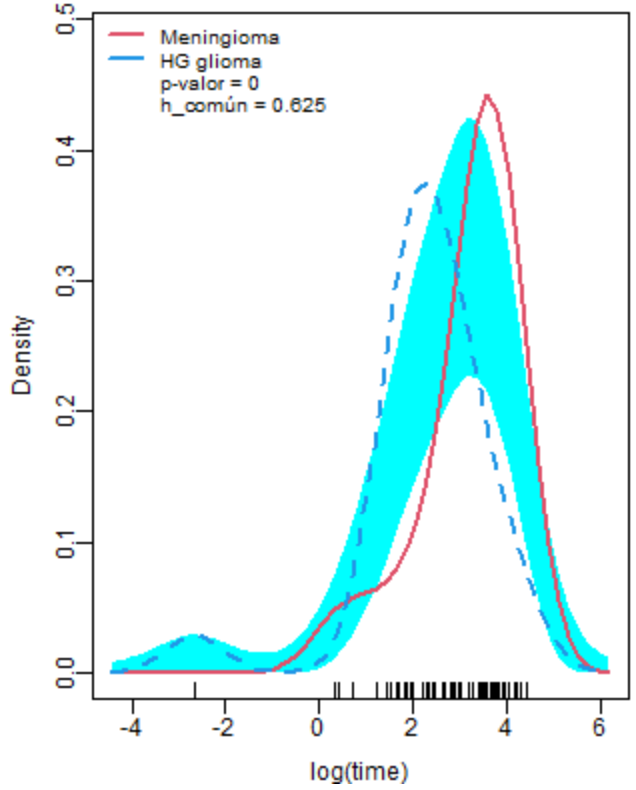
# Aplicación a datos reales

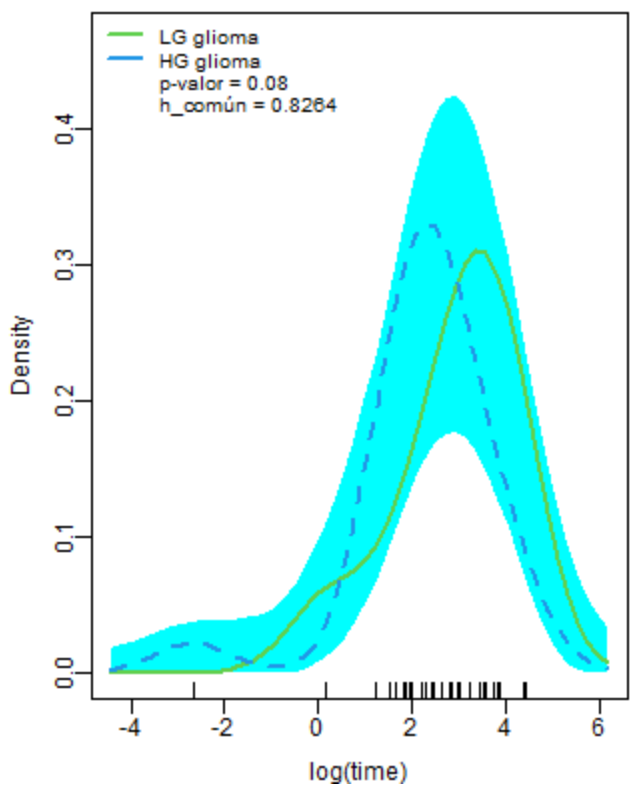
- Conjunto de datos  de Brain Cancer: Meningioma, LG glioma, HG glioma



# Aplicación a datos reales

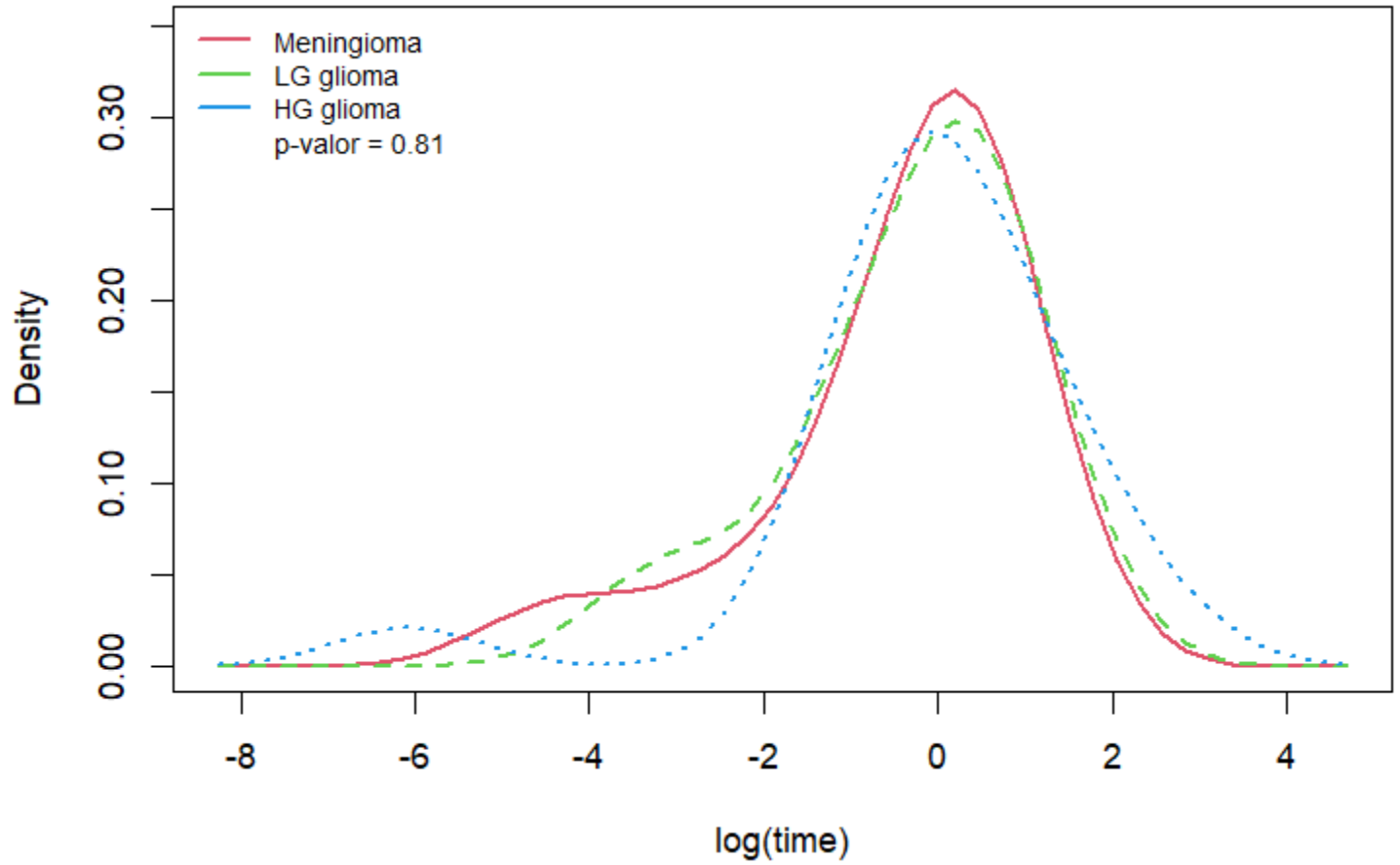


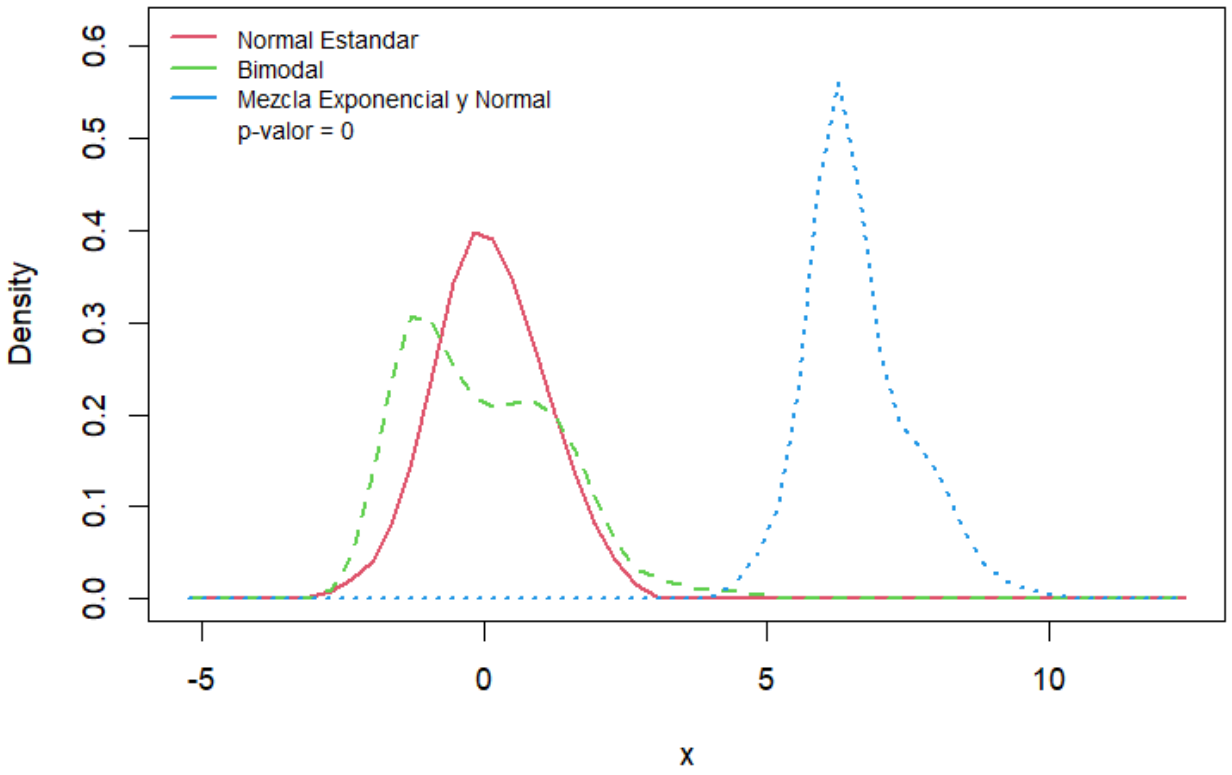
# Aplicación a datos reales

- Estandarizando los tiempos




Diferencia únicamente en localización

# Aplicación a datos simulados




Test No Paramétrico basado en suma de rangos

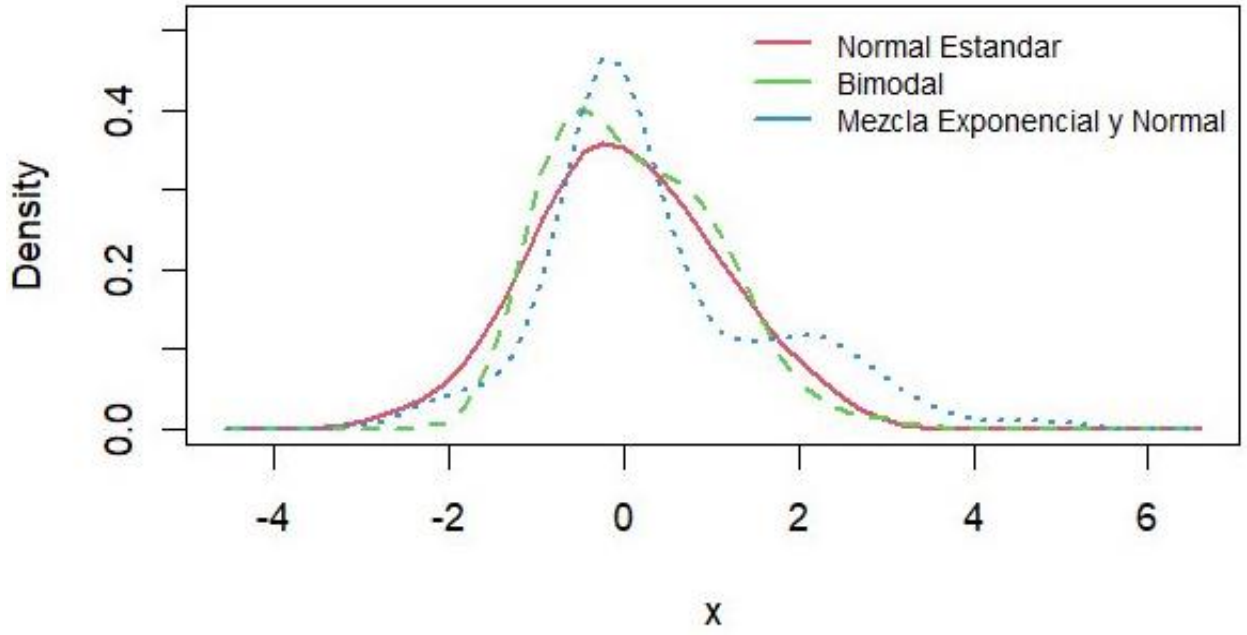
```
        Kruskal-Wallis rank sum test

data:  x by datos$label
Kruskal-Wallis chi-squared = 200.21, df = 2, p-value < 2.2e-16
```

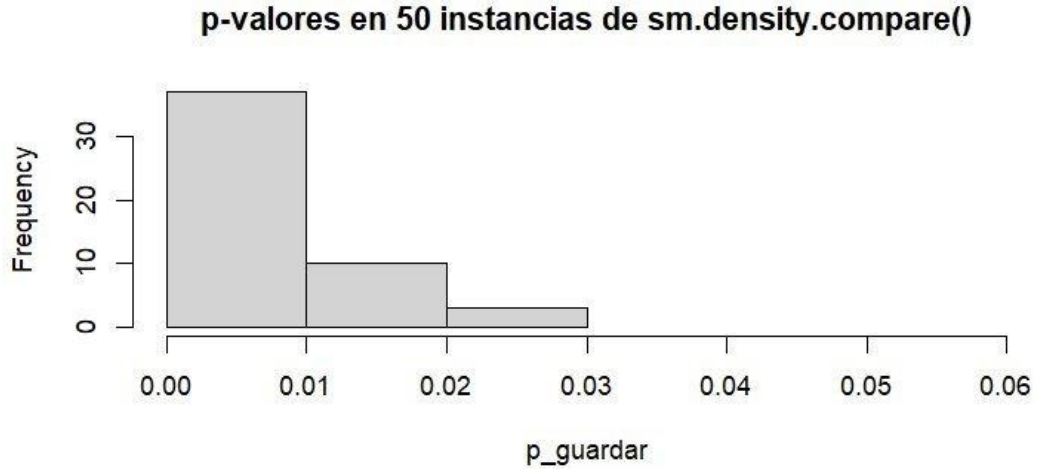
# Aplicación a datos simulados

- Datos estandarizados

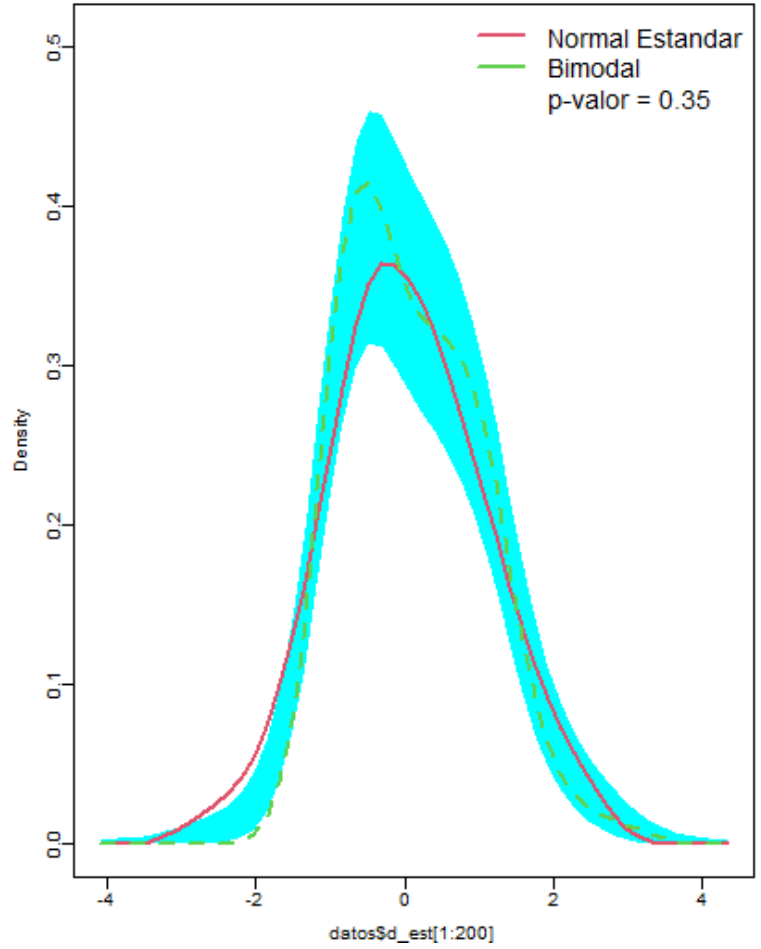
```
>
> kruskal.test(datos$d_est ~ datos$label)

        Kruskal-Wallis rank sum test

data:  datos$d_est by datos$label
Kruskal-Wallis chi-squared = 3.4502, df = 2, p-value = 0.1782

> kwAllPairsNemenyiTest(datos$d_est, datos$label, dist="Chisq")

        Pairwise comparisons using Nemenyi's all-pairs test with chi-square approximation

data: datos$d_est and datos$label

  1    2
2 0.96 -
3 0.22 0.35

P value adjustment method: single-step
alternative hypothesis: two.sided
> pairwise.wilcox.test(datos$d_est, datos$label,
+                      p.adjust.method = "bonf" )

        Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data:  datos$d_est and datos$label

  1    2
2 1.00 -
3 0.29 0.39

P value adjustment method: bonferroni
>
```

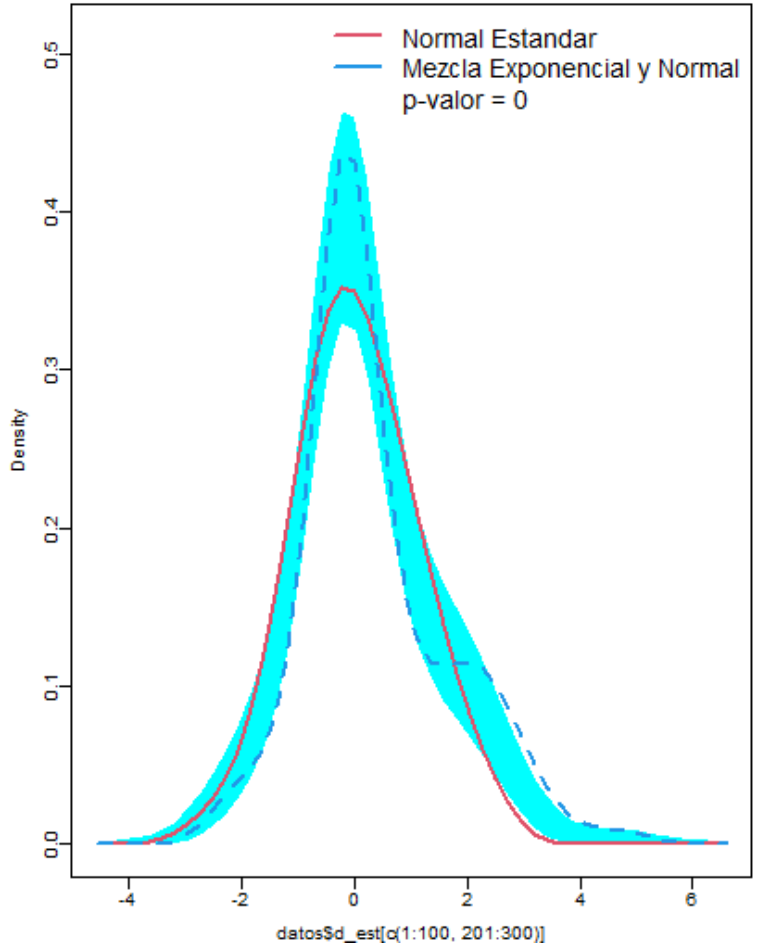
# Aplicación a datos simulados

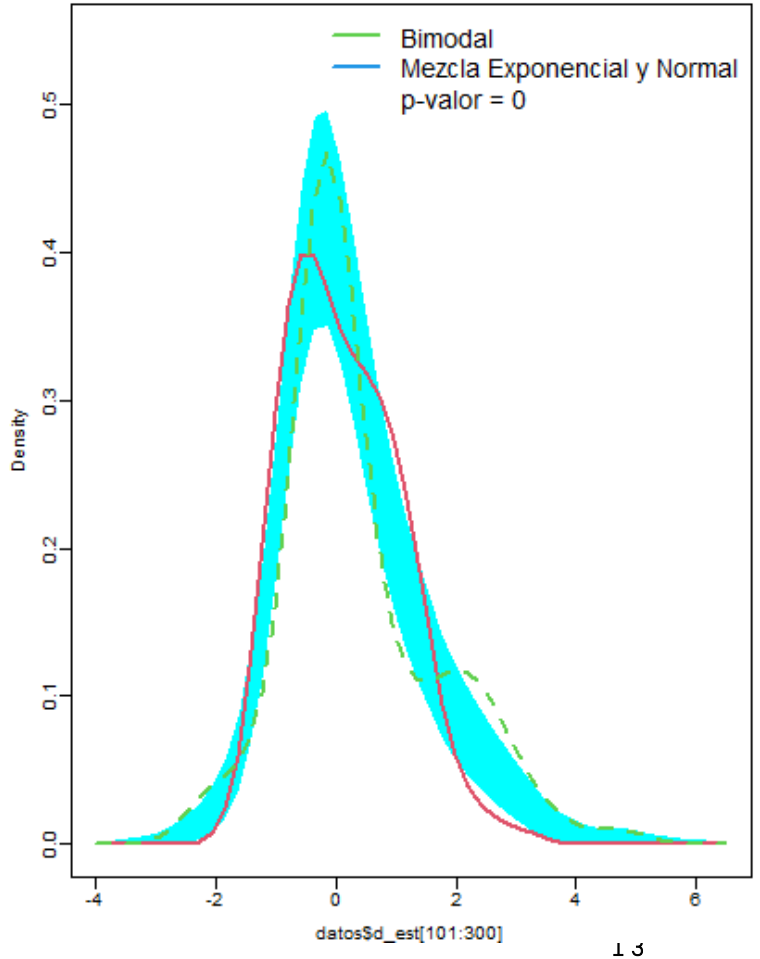


# Aplicación a datos simulados



# Conclusiones

- Comparación distribuciones más allá de diferencias en medias o varianzas.
- sm.density.compare integra visualización, contraste de hipótesis y evaluación gráfica.
- Ancho de banda común, optimalidad del estimador y adaptabilidad a distintos datos.
- Gran valor del suavizado kernel