

[Fecha de última modificación]: 21/03/2026.

Vamos a suponer que en una etapa de REVISIÓN (Paper de clasificación monosensor, Sensors | MDPI) se solicita profundizar el análisis estadístico. Actualmente tenemos una tabla con los valores de F1-score para los 48 experimentos (30 repeticiones en cada experimento), junto con un estudio ANOVA post-hoc de dichas métricas F1.

¿En qué estadísticas podemos ahondar?

Modelo OLS (Ordinary Least Squares, ajuste por mínimos cuadrados)

Sabemos que la relación lineal que hemos escogido para ajustar el modelo ANOVA (4 factores experimentales e interacciones por parejas) es:

F1-score ~ Modelo + Sensor + Config + NumClases + Modelo x Sensor + Modelo x Config + Modelo x NumClases + Sensor x Config + Sensor x NumClases + Config x NumClases.

** Puede ser interesante incorporar todas las interacciones posibles, incluyendo combinaciones de tres factores experimentales. El análisis de todas las interacciones posibles es muy sencillo de implementar en Python: **F1-score** ~ Modelo * Sensor * Config * NumClases.

Comprobaciones previas a ANOVA

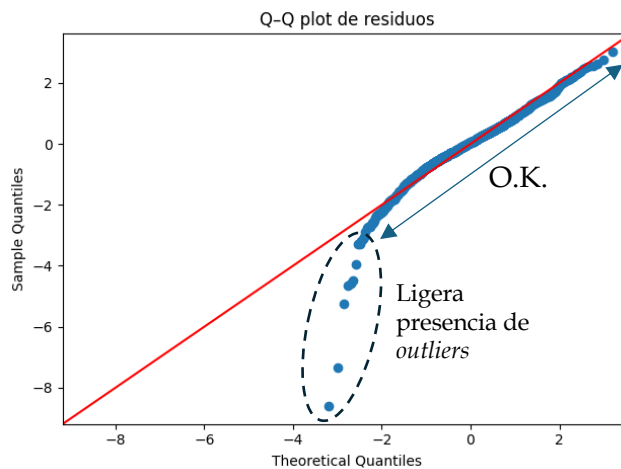
En primer lugar, las asunciones que deberían darse para realizar un ANOVA sobre los valores de F1-score son las siguientes:

- **Normalidad de los residuos** (se comprueba mediante el [test de Shapiro-Wilk](#) y/o la inspección visual de los [gráficos Q-Q](#)).

Nuestro análisis dispone de 48 experimentos con 30 repeticiones cada uno. En total, tenemos $n = 1440$ muestras para realizar el ANOVA.

Salida de Python:

Shapiro-Wilk $W=0.9407547871108994$, $p=1.5065327354014808e-23$



[Fecha de última modificación]: 21/03/2026.

En el gráfico Q-Q se observa que los residuos del ANOVA aplicado sobre los valores de F1-score se aproximan a una distribución normal, si bien existe cierta asimetría de la función de densidad en las colas (pequeño número de muestras que no siguen la tendencia $y=x$ en el Q-Q plot).

El test de Shapiro-Wilk propone aceptar la hipótesis alternativa de no normalidad en los datos, dado que el p-valor obtenido es inferior al nivel de significación considerado ($\alpha=0.05$). No obstante, el elevado tamaño de n agudiza la sensibilidad del test hacia cualquier *outlier* que no cumpla la condición de normalidad; perdiendo su capacidad de generalización. Este hecho, unido a la inspección visual del gráfico Q-Q, sugiere aceptar la distribución normal de los datos; garantizando la robustez de los estadísticos F obtenidos durante el posterior test ANOVA. Además, el estadístico W de valor 0.94 sugiere la existencia de una distribución normal aceptable; dado que $W \in [0.92, 0.96]$.

- Existencia de **homocedasticidad** (homogeneidad entre las varianzas de cada grupo). De esta forma, se evita la presencia de sesgo en el estadístico F de ANOVA. Lo comprobamos mediante el **test de Levene** y su variante **Brown-Forsythe** (empleando medianas).

El **test de Levene/Brown-Forsythe** revela la existencia de heterocedasticidad en los datos, lo cual es coherente con los niveles presentes en los 4 factores experimentales. Por ejemplo, ya observábamos en nuestros boxplots del paper que los modelos convolucionales presentan más variabilidad en su predicción que los decision tree-based. Asimismo, la granularidad de la actividad también introduce modificaciones en la variabilidad de los F1-scores obtenidos.

Salida de Python:

```
,stat,p
Modelo,84.77192523552051,1.6870091902450846e-50
Sensor,43.130441979164125,6.447043493421842e-19
Config,64.68946113276675,1.8163287236471823e-15
NumClases,64.81764967154744,1.7067074954576558e-15
*****
```

Al igual que sucede con Shapiro-Wilk, se rechaza la hipótesis nula del test, lo que sugeriría la ausencia de homocedasticidad entre las varianzas de cada factor experimental. Todos los p-valores obtenidos son menores que α , por lo que Levene sugiere aceptar H_1 . No obstante, la explicación relativa a los Q-Q plots sigue siendo aplicable en este estudio de homocedasticidad; validando la fiabilidad estadística del ANOVA post-hoc.

- **Observaciones estadísticamente independientes.** No existe un test estadístico para comprobarlo, sino que está ligado al diseño experimental.

El protocolo de captura de datos garantizaría la independencia estadística de las muestras poblacionales observadas.

Análisis posteriores a ANOVA:

○ **Medias Marginales Estimadas (EMMs - Estimated Marginal Means)**

Las **EMMs** nos permiten **comparar de forma aislada el efecto** de los diferentes niveles dentro de cada uno de los 4 factores (modelo, localización del sensor, configuración del sensor y granularidad de la actividad física). Por tanto, las EMMs de un factor proporcionan las medias de F1-score **habiendo eliminado las distorsiones que pudieran provocar el resto de factores** (las EMMs se obtienen a partir del modelo ANOVA).

Salida de Python:

```
/ EMM_Config
  Acc,0.6539383333333301
  Acc+Gyro,0.736487777777717
/ EMM_Modelo
  Random Forest,0.7339411111111049
  XGBoost,0.733224722222169
  CNN ESANN,0.619600555555506
  CNN CAPTURE-24,0.694085833333315
/ EMM_NumClases
  15,0.626029166666661
  4,0.764396944444408
/ EMM_Sensor
  T,0.708851874999975
  W,0.725194374999936
  H,0.651592916666619
```

Se corroboran las principales conclusiones a las que habíamos llegado en el estudio previo a la revisión estadística: (1) la incorporación del giroscopio produce un incremento estadísticamente significativo del F1-score, (2) los modelos decisión tree-based ofrecen mayor rendimiento de clasificación que los convolucionales, (3) disminuir la granularidad de la actividad física incrementa significativamente el valor de F1, y (4) los resultados apuntan que la muñeca es la localización corporal óptima del IMU.

○ **Test de Tukey HSD (Honest Significant Difference)**

El test ANOVA nos reveló previamente que existen diferencias entre los niveles de cada factor. Sin embargo, no especifica **qué nivel es el óptimo** (aunque ya lo hemos intuido visualmente mediante nuestros boxplots), **qué amplitud tiene cada diferencia**, la **significatividad** de la diferencia o **qué nivel supera a otro**. Estas respuestas nos las proporciona numéricamente el test de Tukey, el cual complementa (o incluso sustituye) al test ANOVA previo. La esencia del test de Tukey está en el cálculo de la Diferencia Mínima Honesta (HSD), tenemos una expresión analítica para su cálculo aunque ya está todo el test implementado en librerías de Python.

[Fecha de última modificación]: 21/03/2026.

Además de comparar los niveles dentro un mismo factor, el test de Tukey también puede calcular los efectos simples de un factor dentro de cada nivel de otro factor. Algunos ejemplos: comparar modelos dentro de una localización de sensor, comparar granularidad de actividad dentro un tipo de modelo, etc. Puede ser interesante y metodológicamente más exhaustivo que solo el ANOVA.

El contraste de hipótesis del test de Tukey es:

$H_0: \mu_0 = \mu_1$ (no hay diferencia estadísticamente significativa entre las medias poblacionales)

$H_1: \mu_0 \neq \mu_1$ (las medias poblacionales son estadísticamente diferentes)

Vamos a calcularlo:

Salida de Python (+ nuestra interpretación de los resultados):

```
group1,group2,meandiff,p-adj,lower,upper,reject,Factor,Nivel
//
CNN CAPTURE-24,CNN ESANN,-0.0745,0.0,-0.0958,-0.0531,True,Modelo,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los 2 niveles convolucionales del factor Modelo).
//
CNN CAPTURE-24,Random Forest,0.0399,0.0,0.0185,0.0612,True,Modelo,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y RF del factor Modelo).
//
CNN CAPTURE-24,XGBoost,0.0391,0.0,0.0178,0.0605,True,Modelo,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y XGB del factor Modelo).
//
CNN ESANN,Random Forest,0.1143,0.0,0.093,0.1357,True,Modelo,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y RF del factor Modelo).
//
CNN ESANN,XGBoost,0.1136,0.0,0.0923,0.135,True,Modelo,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGB del factor Modelo).
//
Random Forest,XGBoost,-0.0007,0.9998,-0.0221,0.0206,False,Modelo,GLOBAL (∄ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGB del factor Modelo).
//
H,T,0.0573,0.0,0.0396,0.0749,True,Sensor,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Thigh del factor Sensor).
//
H,W,0.0736,0.0,0.056,0.0912,True,Sensor,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Wrist del factor Sensor).
//
T,W,0.0163,0.0763,-0.0013,0.034,False,Sensor,GLOBAL (∄ diferencias estadísticamente significativas entre las medias de F1-score para los niveles Thigh y Wrist del factor Sensor).
//
Acc,Acc+Gyro,0.0825,0.0,0.0708,0.0943,True,Config,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles Acc y Acc+Gyr del factor Config).
//
4,15,-0.1384,0.0,-0.1486,-0.1281,True,NumClases,GLOBAL (∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles 4 y 15 del factor NumClases).
//
// A continuación, analizamos la influencia de las interacciones por parejas de factores:
```

[Fecha de última modificación]: 21/03/2026.

```
//  
CNN CAPTURE-24,CNN ESANN,-0.1368,0.0,-0.178,-0.0956,True,Modelo,Sensor=T (Dentro del nivel  
Thigh del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).  
//  
CNN CAPTURE-24,Random Forest,-0.0221,0.5093,-0.0634,0.0191,False,Modelo,Sensor=T (Dentro  
del nivel Thigh del factor Sensor, ∄ diferencias estadísticamente significativas entre las medias de F1-score  
para los niveles CNN CAPTURE-24 y Random Forest del factor Modelo).  
//  
CNN CAPTURE-24,XGBoost,-0.0193,0.6216,-0.0605,0.0219,False,Modelo,Sensor=T (Dentro del nivel  
Thigh del factor Sensor, ∄ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles CNN CAPTURE-24 y XGBoost del factor Modelo).  
//  
CNN ESANN,Random Forest,0.1146,0.0,0.0734,0.1558,True,Modelo,Sensor=T (Dentro del nivel  
Thigh del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles CNN ESANN y Random Forest del factor Modelo).  
//  
CNN ESANN,XGBoost,0.1175,0.0,0.0762,0.1587,True,Modelo,Sensor=T (Dentro del nivel Thigh del  
factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN  
ESANN y XGBoost del factor Modelo).  
//  
Random Forest,XGBoost,0.0028,0.998,-0.0384,0.044,False,Modelo,Sensor=T (Dentro del nivel  
Thigh del factor Sensor, ∄ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles Random Forest y XGBoost del factor Modelo).  
//  
CNN CAPTURE-24,CNN ESANN,-0.0554,0.0,-0.0823,-0.0285,True,Modelo,Sensor=W (Dentro del nivel  
Wrist del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).  
//  
CNN CAPTURE-24,Random Forest,0.0512,0.0,0.0242,0.0781,True,Modelo,Sensor=W (Dentro del nivel  
Wrist del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles CNN CAPTURE-24 y Random Forest del factor Modelo).  
//  
CNN CAPTURE-24,XGBoost,0.0543,0.0,0.0274,0.0812,True,Modelo,Sensor=W (Dentro del nivel Wrist  
del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles  
CNN CAPTURE-24 y XGBoost del factor Modelo).  
//  
CNN ESANN,Random Forest,0.1065,0.0,0.0796,0.1335,True,Modelo,Sensor=W (Dentro del nivel Wrist  
del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles  
CNN ESANN y Random Forest del factor Modelo).  
//  
CNN ESANN,XGBoost,0.1097,0.0,0.0828,0.1366,True,Modelo,Sensor=W (Dentro del nivel Wrist del  
factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN  
ESANN y XGBoost del factor Modelo).  
//  
Random Forest,XGBoost,0.0032,0.9902,-0.0237,0.0301,False,Modelo,Sensor=W (Dentro del nivel  
Wrist del factor Sensor, ∄ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles Random Forest y XGBoost del factor Modelo).  
//  
CNN CAPTURE-24,CNN ESANN,-0.0313,0.1069,-0.0668,0.0043,False,Modelo,Sensor=H (Dentro del  
nivel Hip del factor Sensor, ∄ diferencias estadísticamente significativas entre las medias de F1-score para  
los niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).  
//  
CNN CAPTURE-24,Random Forest,0.0906,0.0,0.055,0.1261,True,Modelo,Sensor=H (Dentro del nivel  
Hip del factor Sensor, ∃ diferencias estadísticamente significativas entre las medias de F1-score para los  
niveles CNN CAPTURE-24 y Random Forest del factor Modelo).  
//
```

[Fecha de última modificación]: 21/03/2026.

CNN CAPTURE-24,XGBoost,0.0824,**0.0**,0.0469,0.118,**True**,Modelo,Sensor=H (Dentro del nivel Hip del factor Sensor, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y XGBoost del factor Modelo).

//

CNN ESANN,Random Forest,0.1218,**0.0**,0.0863,0.1574,**True**,Modelo,Sensor=H (Dentro del nivel Hip del factor Sensor, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y Random Forest del factor Modelo).

//

CNN ESANN,XGBoost,0.1137,**0.0**,0.0781,0.1492,**True**,Modelo,Sensor=H (Dentro del nivel Hip del factor Sensor, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGBoost del factor Modelo).

//

Random Forest,XGBoost,-0.0081,**0.9349**,-0.0437,0.0274,**False**,Modelo,Sensor=H (Dentro del nivel Hip del factor Sensor, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Random Forest y XGBoost del factor Modelo).

CNN CAPTURE-24,CNN ESANN,-0.0698,**0.0**,-0.0995,-0.0401,**True**,Modelo,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).

//

CNN CAPTURE-24,Random Forest,0.0825,**0.0**,0.0528,0.1122,**True**,Modelo,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y Random Forest del factor Modelo).

//

CNN CAPTURE-24,XGBoost,0.082,**0.0**,0.0523,0.1117,**True**,Modelo,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y XGBoost del factor Modelo).

//

CNN ESANN,Random Forest,0.1523,**0.0**,0.1226,0.182,**True**,Modelo,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y Random Forest del factor Modelo).

//

CNN ESANN,XGBoost,0.1518,**0.0**,0.1221,0.1815,**True**,Modelo,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGBoost del factor Modelo).

//

Random Forest,XGBoost,-0.0005,**1.0**,-0.0302,0.0292,**False**,Modelo,Config=Acc (Dentro del nivel Acc del factor Config, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Random Forest y XGBoost del factor Modelo).

//

CNN CAPTURE-24,CNN ESANN,-0.0792,**0.0**,-0.1044,-0.0539,**True**,Modelo,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).

//

CNN CAPTURE-24,Random Forest,-0.0028,**0.9922**,-0.028,0.0225,**False**,Modelo,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y Random Forest del factor Modelo).

//

CNN CAPTURE-24,XGBoost,-0.0037,**0.9817**,-0.029,0.0216,**False**,Modelo,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y XGBoost del factor Modelo).

//

CNN ESANN,Random Forest,0.0764,**0.0**,0.0511,0.1016,**True**,Modelo,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y Random Forest del factor Modelo).

//

[Fecha de última modificación]: 21/03/2026.

CNN ESANN,XGBoost,0.0755,**0.0**,0.0502,0.1007,**True**,Modelo,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGBoost del factor Modelo).

//

Random Forest,XGBoost,-0.0009,**0.9997**,-0.0262,0.0243,**False**,Modelo,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Random Forest y XGBoost del factor Modelo).

//

CNN CAPTURE-24,CNN ESANN,-0.126,**0.0**,-0.1506,-0.1013,**True**,Modelo,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).

//

CNN CAPTURE-24,Random Forest,0.0344,**0.0019**,0.0098,0.0591,**True**,Modelo,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).

//

CNN CAPTURE-24,XGBoost,0.0217,**0.1074**,-0.003,0.0463,**False**,Modelo,NumClases=15 (Dentro del nivel 15 del factor NumClases, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y XGBoost del factor Modelo).

//

CNN ESANN,Random Forest,0.1604,**0.0**,0.1358,0.185,**True**,Modelo,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y Random Forest del factor Modelo).

//

CNN ESANN,XGBoost,0.1476,**0.0**,0.123,0.1722,**True**,Modelo,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGBoost del factor Modelo).

//

Random Forest,XGBoost,-0.0128,**0.5408**,-0.0374,0.0119,**False**,Modelo,NumClases=15 (Dentro del nivel 15 del factor NumClases, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Random Forest y XGBoost del factor Modelo).

//

CNN CAPTURE-24,CNN ESANN,-0.023,**0.0297**,-0.0445,-0.0016,**True**,Modelo,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y CNN ESANN del factor Modelo).

//

CNN CAPTURE-24,Random Forest,0.0453,**0.0**,0.0238,0.0667,**True**,Modelo,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y Random Forest del factor Modelo).

//

CNN CAPTURE-24,XGBoost,0.0566,**0.0**,0.0352,0.0781,**True**,Modelo,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN CAPTURE-24 y XGBoost del factor Modelo).

//

CNN ESANN,Random Forest,0.0683,**0.0**,0.0469,0.0897,**True**,Modelo,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y Random Forest del factor Modelo).

//

CNN ESANN,XGBoost,0.0796,**0.0**,0.0582,0.1011,**True**,Modelo,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles CNN ESANN y XGBoost del factor Modelo).

//

Random Forest,XGBoost,0.0113,**0.5237**,-0.0101,0.0328,**False**,Modelo,NumClases=4 (Dentro del nivel 4 del factor NumClases, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Random Forest y XGBoost del factor Modelo).

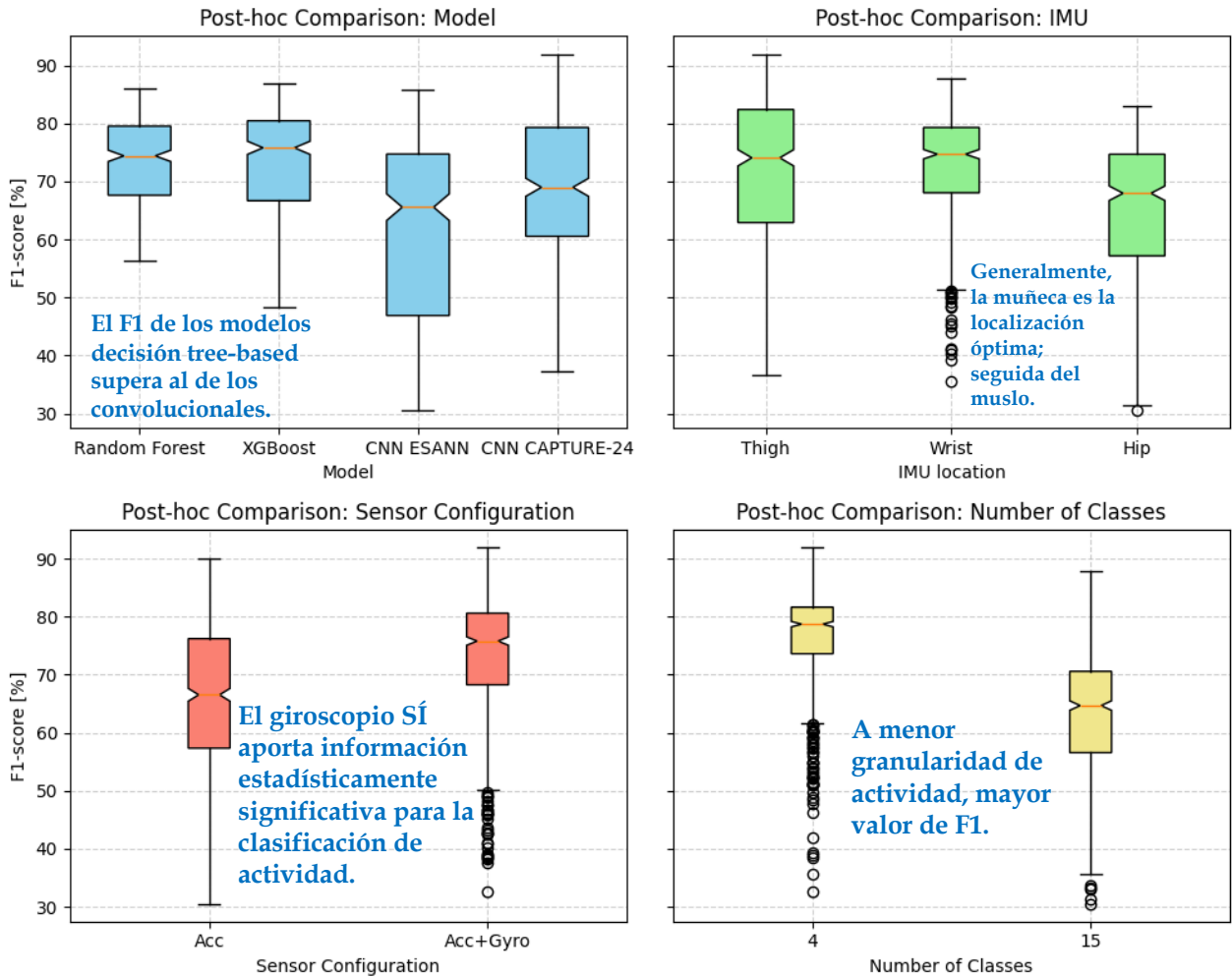
//

[Fecha de última modificación]: 21/03/2026.

H,T,0.0951,**0.0**,0.0695,0.1207,**True**,Sensor,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Thigh del factor Sensor).
//
H,W,0.0821,**0.0**,0.0565,0.1077,**True**,Sensor,Config=Acc (Dentro del nivel Acc del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Wrist del factor Sensor).
//
T,W,-0.013,**0.4562**,-0.0386,0.0126,**False**,Sensor,Config=Acc (Dentro del nivel Acc del factor Config, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Thigh y Wrist del factor Sensor).
//
H,T,0.0194,**0.0656**,-0.001,0.0398,**False**,Sensor,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Thigh del factor Sensor).
//
H,W,0.0651,**0.0**,0.0448,0.0855,**True**,Sensor,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Wrist del factor Sensor).
//
T,W,0.0457,**0.0**,0.0254,0.0661,**True**,Sensor,Config=Acc+Gyro (Dentro del nivel Acc+Gyro del factor Config, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Thigh y Wrist del factor Sensor).
//
H,T,0.0081,**0.6529**,-0.0136,0.0298,**False**,Sensor,NumClases=15 (Dentro del nivel 15 del factor NumClases, \nexists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Thigh del factor Sensor).
//
H,W,0.1003,**0.0**,0.0786,0.122,**True**,Sensor,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Wrist del factor Sensor).
//
T,W,0.0922,**0.0**,0.0705,0.1139,**True**,Sensor,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Thigh y Wrist del factor Sensor).
//
H,T,0.1064,**0.0**,0.0907,0.1221,**True**,Sensor,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Thigh del factor Sensor).
//
H,W,0.0469,**0.0**,0.0311,0.0626,**True**,Sensor,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Hip y Wrist del factor Sensor).
//
T,W,-0.0595,**0.0**,-0.0753,-0.0438,**True**,Sensor,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Thigh y Wrist del factor Sensor).
//
Acc,Acc+Gyro,0.0978,**0.0**,0.0832,0.1123,**True**,Config,NumClases=15 (Dentro del nivel 15 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Acc y Acc+Gyro del factor Config).
//
Acc,Acc+Gyro,0.0673,**0.0**,0.0559,0.0788,**True**,Config,NumClases=4 (Dentro del nivel 4 del factor NumClases, \exists diferencias estadísticamente significativas entre las medias de F1-score para los niveles Acc y Acc+Gyro del factor Config).

[Fecha de última modificación]: 21/03/2026.

Observamos que tanto el test de Tukey como la estimación de medias marginales (EMMs) nos corroboran estadísticamente las evidencias que habíamos intuido previamente mediante la inspección visual de nuestros boxplots:



El código Python que hemos implementado para obtener los resultados estadísticos está en el repo Git de ML:

[uniovi-simur-wearablepermed-ml/src/wearablepermed_ml/Advanced_Statistical_analysis/](https://github.com/uniovi-simur-wearablepermed-ml/src/wearablepermed_ml/Advanced_Statistical_analysis/)