

Statistic Test

En esta practica aparte de la librería de pandas utilice las librerías scipy y statsmodels para realizar las pruebas elegi los datos etiquetados “Category” y “Region”. Realice solamente las pruebas ANOVA Y KRUSKAL-WALLIS y adicional realice la prueba TUKEY HDS para comparar los grupos en pares.

```
ANOVA: F=262.165, p=0.0000
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1      group2      meandiff p-adj  lower  upper  reject
-----
Furniture Office Supplies -231.2728  0.0 -267.7759 -194.7697  True
Furniture Technology 105.7477  0.0  59.751 151.7444  True
Office Supplies Technology 337.0205  0.0 298.5943 375.4466  True
-----
Kruskal-Wallis: H=25.247, p=0.0000
```

En la prueba ANOVA para el grupo “Category”, el valor de **F=262.165** con un **p-valor=0.0000** indica que existen diferencias estadísticamente significativas entre las medias de al menos dos de los grupos analizados (Furniture, Office Supplies, Technology). El p-valor extremadamente bajo (menor que 0.05) confirma que estas diferencias no son debidas al azar.

Como el p-valor es significativo se realizo la prueba TUKEY HSD y se puede concluir que todos los pares de grupos muestran diferencias significativas (p-adj=0.0 en todos los casos), ya que el valor "reject" es True para cada comparación.

- **Furniture vs. Office Supplies:** La diferencia media es **-231.2728**, lo que indica que las ventas/valores de Furniture son significativamente menores que los de Office Supplies.
- **Furniture vs. Technology:** La diferencia media es **105.7477**, lo que indica que las ventas/valores de Furniture son significativamente mayores que los de Technology.
- **Office Supplies vs. Technology:** La diferencia media es **337.0205**, lo que indica que las ventas/valores de Office Supplies son significativamente mayores que los de Technology.

Y la última prueba para las regiones también confirma que hay diferencias significativas entre los grupos, con un valor de **H=25.247** y un **p-valor=0.0000**. Esto respalda los resultados del ANOVA, indicando que las diferencias no se deben a la distribución de los datos.