

Tema 2

Estadística robusta

Técnicas multivariantes

Dr. Antoni Ferragut



Estadística clásica vs Estadística robusta

Supuestos

- Normalidad
- Homocedasticidad
- Independencia de los datos
- Tamaño de la muestra

Outliers

- Obviados o no se cumplen supuestos \Rightarrow pérdida de fiabilidad

Estimadores

- Localización: media
- Dispersión: desviación típica

No deberían usarse cuando

- No se cumplen los supuestos del modelo
- Se sospecha de la existencia de outliers
- La muestra no procede de una sola población

Se sigue una distribución normal

Pruebas:

- Kolmogorov-Smirnov ($n \geq 5000$)
- Shapiro-Wilk ($n \leq 5000$)
- Técnicas de visualización \Rightarrow outliers, errores

- La varianza de los errores es constante a lo largo del tiempo
- Heterocedasticidad \rightarrow valores atípicos o con demasiado peso

Pruebas:

- Visualización gráfica \Rightarrow dispersión de los residuos
- Breusch-Pagan

- Necesitamos un tamaño muestral mínimo
- Tamaño elevado \Rightarrow TCL

- Útiles cuando alguno de los supuestos anteriores no se cumplen

Limitaciones:

- Menos potentes
- Requieren modificación de hipótesis

Pruebas paramétricas y no paramétricas		
Tipo de prueba	Paramétrica	No Paramétrica
Describir una muestra	Media y varianza	Mediana, rango intercuartil
Comparar una muestra con un valor	Test t de una muestra (media)	Prueba Wilcoxon (mediana)
Comparar dos muestras independientes	Test t de igualdad de medias	Prueba Mann-Whitney (mediana)
Comparar dos muestras apareadas	Test t de igualdad de medias apareadas	Prueba Wilcoxon (mediana)
Para n muestras	ANOVA	Prueba Kruskal-Wallis

- Muy comunes
- Pueden afectar a la estimación de los parámetros

Opciones:

- Detectar y eliminar (perdemos información y tamaño muestral; ok para errores)
- Diseñar métodos robustos no muy afectados por ellos (menos eficientes si no los hay)

Opciones:

- Mediana
- (α) -Mediana recortada
- Media winsorizada
- Trimedia

Observaciones:

- En todos los casos perdemos eficiencia
- Más robustos cuando hay valores anómalos
- Podemos comprobar si hay outliers usando estadísticos clásicos y robustos y comparar
- Robustos y eficientes: M-estimadores (Huber) \Rightarrow software

Opciones:

- MAD (desviación absoluta mediana)
- Rango intercuartílico
- M-estimador de un paso