Universidad **Ricardo Palma**

RECTORADO

PROGRAMA DE ESPECIALIZACIÓN EN CIENCIA DE DATOS

Formamos seres humanos para una cultura de paz

TALLER DE ESTADÍSTICA PARA LA CIENCIA DE DATOS

R + Python

MÓDULO 5 : ANALISIS MULTIVARIADO II

Análisis de Varianza - ANOVA

A nuestro recordado Maestro

Dr. Erwin Kraenau Espinal, Presidente de la Comisión de Creación de la Maestría en Ciencia de los Datos

TALLER DE ESTADÍSTICA PARA CIENCIA DE DATOS

EXPOSITORES

José Antonio Cárdenas Garro
UNMSM
MSc in Data Science Candidate
Promotion "Erwin Kraenau Espinal"
Universidad Ricardo Palma

Predictive Modelling Specialist — Scotiabank

Portfolio and Consumption Analyst — Scotiabank

Data Analyst — ONP (Oficina de Normalización Previsional)

Correo : josecardenasgarro@gmail.com
LinkedIn : www.linkedin.com/in/jos%C3%A9-antonio-c%C3%A1rdenas-garro-599266b0

André Omar Chávez Panduro
UNMSM
MSc in Data Science Candidate
Promotion "Erwin Kraenau Espinal"
Universidad Ricardo Palma

Data Scientist — Interbank

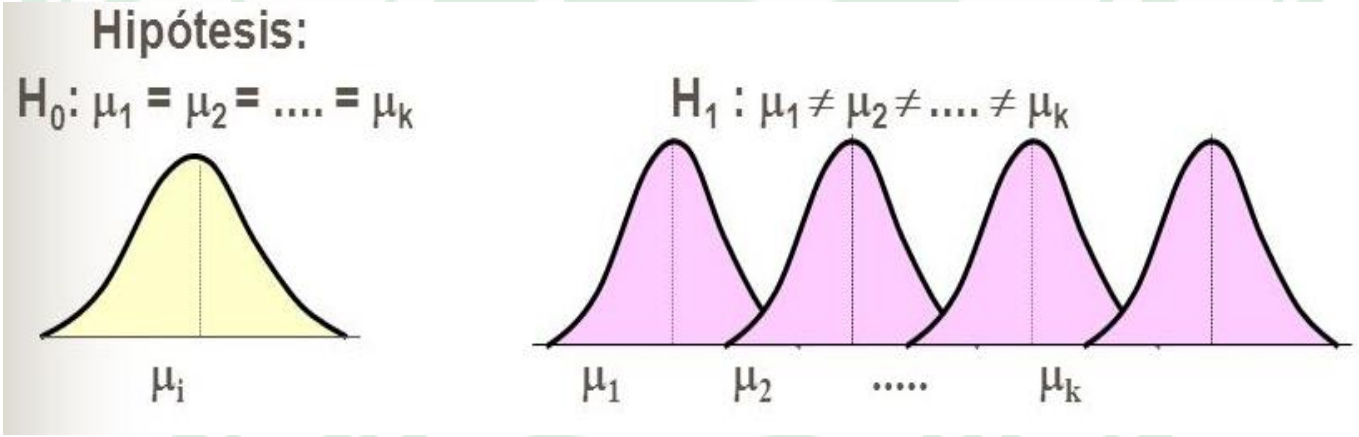
Customer Intelligence Analyst — Claro

Data Analyst — INEI (Instituto Nacional de Estadística e Informática)

Correo : andrecp38@gmail.com / 09140205@unmsm.edu.pe
LinkedIn : www.linkedin.com/in/andré-chávez-a90078b9

« Divide las dificultades que examinas en tantas partes como sea posible , para su mejor solución»

Análisis de Varianza. ANOVA



ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

PROBLEMA 1: Dada una variable cuantitativa continua Y, y una variable cualitativa F, determínese si entre ambas hay relación, o no.

Ejemplos: Tiempo de cura / medicamento utilizado
Rendimiento de cosechas / fertilizante
Renta familiar / hábito de lectura
Número de préstamos / ubicación

PROBLEMA 2: Dada una variable cuantitativa continua Y, y varias variables cualitativas F_1, F_2,…, F_n, determínese cuáles de ellas infuyen en Y, y cuáles no (es decir, cuáles guardan relación con Y).

Ejemplos: Tiempo de cura / medicamento utilizado, grupo sanguíneo
Número de préstamos / sexo, nivel de estudios, afición al cine

ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

PROBLEMA 1: Dada una variable cuantitativa continua Y, y una variable cualitativa F, determínese si entre ambas hay relación, o no.

↓

ANOVA simple

Y: variable respuesta (numérica)
F: factor (cualitativa)

PROBLEMA 2: Dada una variable cuantitativa continua Y, y varias variables cualitativas F_1, F_2,..., F_n, determínese cuáles de ellas infuyen en Y, y cuáles no (es decir, cuáles guardan relación con Y).

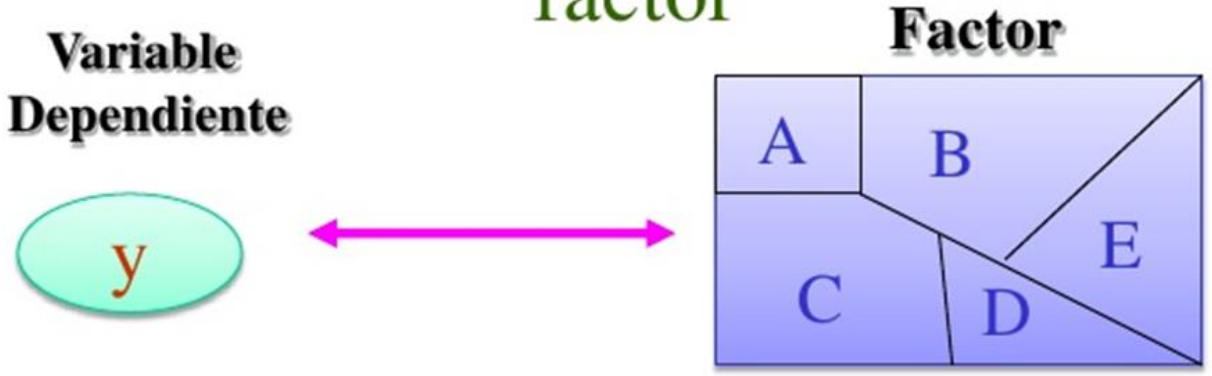
↓

ANOVA multifactorial

Y: variable respuesta (numérica)
F_1, F_2,..., F_n : factores (cualitativas)

ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

Análisis de Varianza de un solo factor



$$y_{ij} = \mu + \tau_i + e_{ij}$$

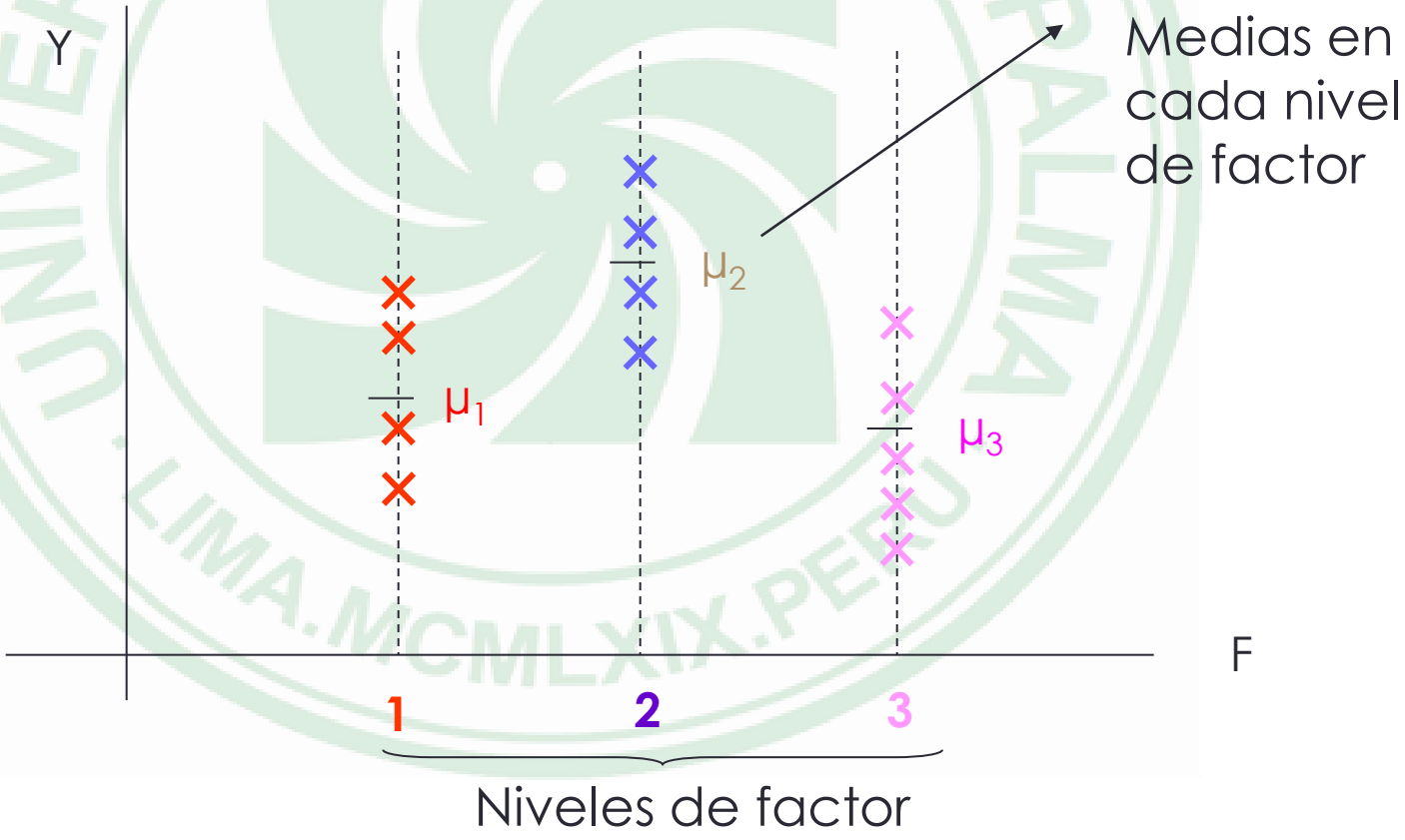
- ✓ e_{ij} *error aleatorio Normales*
- ✓ *Independientes*
- ✓ *Media de cero*
- ✓ *Varianza constante*

ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

1. ANOVA SIMPLE:

Y: variable respuesta (numérica)
F: factor (cualitativa)

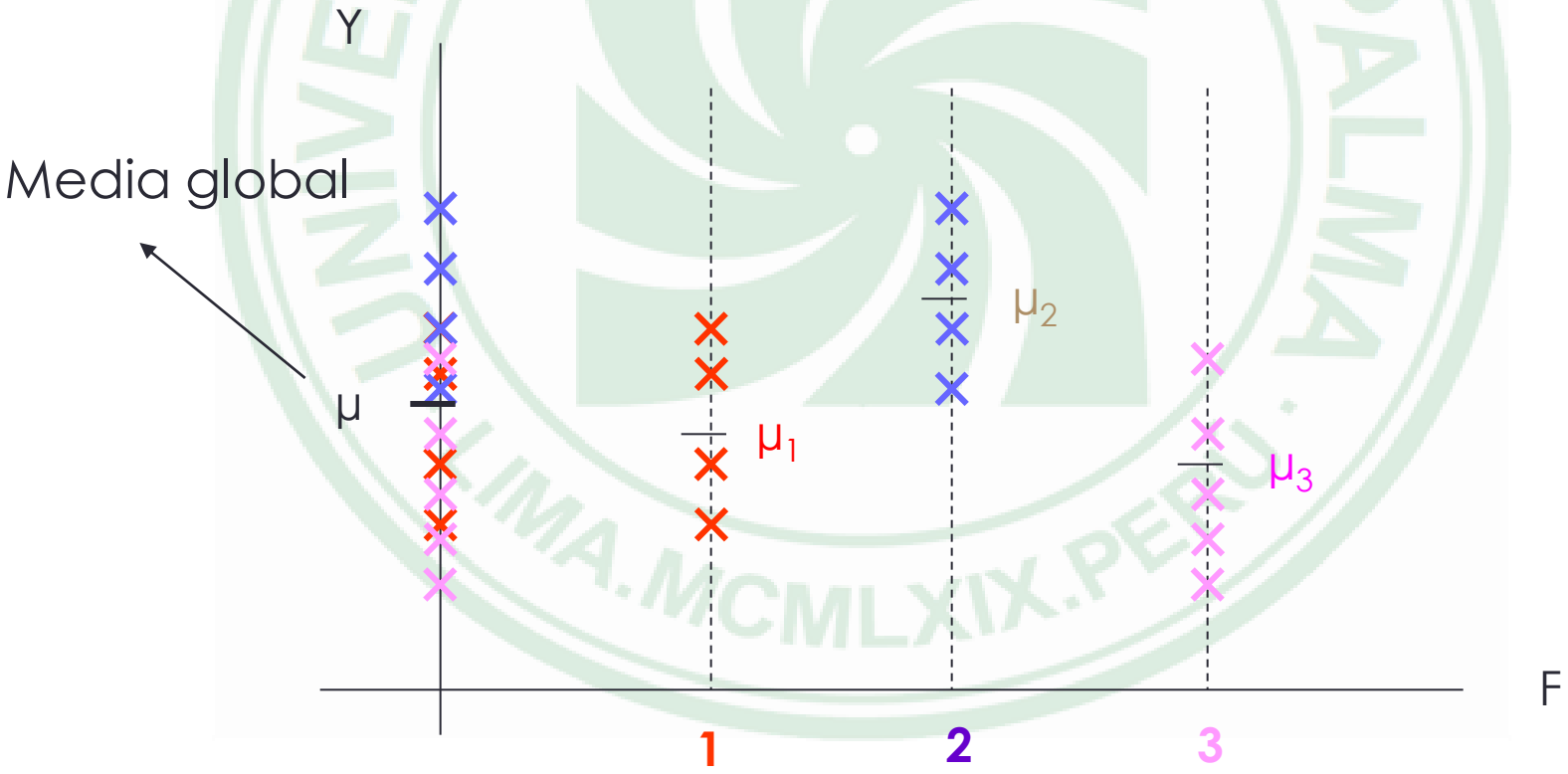
¿Son independientes Y y F? ¿Hay relación entre Y y F? ¿Hay diferencias significativas en el valor de Y, según que F tome uno u otro valor? ¿Influye F en el valor de Y? ¿Hay diferencias en los valores de Y, entre los distintos grupos determinados por F?



ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

Y: variable respuesta (numérica)
F: factor (cualitativa)

Si el valor de F no guarda relación con el de Y… ¿Cómo deberían ser μ_1, μ_2, μ_3?



ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: alguna μ_i es distinta

H_0 equivalente a: Y, F son independientes; Y, F no guardan relación; F no influye en el valor de Y; no hay diferencias significativas en Y según distintos valores de F, etc.

Rechazar H_0 equivale a encontrar dependencia entre F e Y.

ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

¿Cómo contrastar
H_0: $\mu_1 = \mu_2 = \ldots = \mu_n$
H_1: alguna μ_i es distinta ?

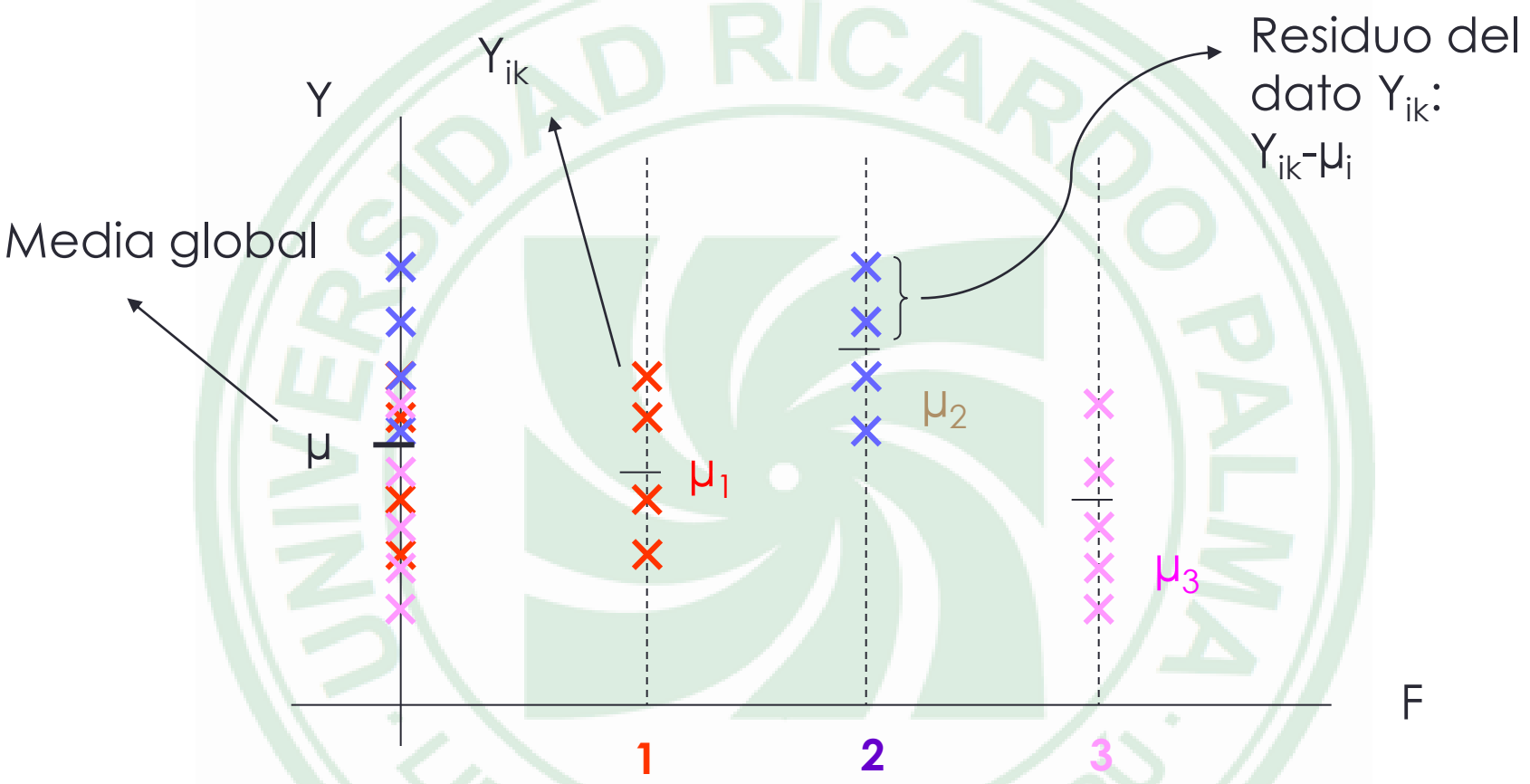
Mala idea: varios contrastes
H_0: $\mu_i = \mu_k$
H_1: $\mu_i \neq \mu_k$

Error de tipo I se acumula, la confianza "total" es demasiado baja

Buena idea: descomposición de la **variabilidad**

Descomposición Variabilidad

ANALISIS DE VARIANZA: IDEA INTUITIVA DE DIFERENCIA DE MEDIAS



Y_{ik}: el primer subíndice (i) indica el valor del nivel del factor; el segundo (k), el orden que ocupa el dato dentro de los pertenecientes a ese nivel del factor.

ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

TABLA DE ANOVA:

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianzas ó cuadrados medios	Cociente-F
Entre-grupos(VE)	$\sum_{i,j}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2$	I-1	$s_e^2 = VE/I-1$	$\hat{s}_e^2 \,/\, \hat{s}_R^2$
Intra-gruposó residual ó no explicada (VNE)	$\sum_{i,j}(y_{ij}-\bar{y}_{i\bullet})^2$	N-I	$s_R^2 = VNE/N-I$	
Total (VT)	$\sum_{i,j}(y_{ij}-\bar{y}_{\bullet\bullet})^2$	N-1		

$$\hat{s}_e^2 \,/\, \hat{s}_R^2 = F_{I-1,N-1}$$

Raíz cuadrada de s_e^2: **error experimental**

TABLA DE ANOVA:

SCE: suma de cuadr. explicada o entre-grupos

Análisis de la Varianza

Fuente	Sumas de cuad.	Gl	Cuadrado Medio	Cociente-F	P
Entre grupos	1,05061E9	3	3,50202E8	1,21	0
Intra grupos	2,69068E10	93	2,8932E8		
Total (Corr.)	2,79574E10	96			

SCR: suma de cuadr. residual o intra-grupos

SCT: suma de cuadr. totales

$$\frac{SCE}{SCT} \times 100 = \text{VARIABILID AD EXPLICADA}$$

ANALISIS DE VARIANZA: ANOVA DE UN FACTOR

IMPORTANTE: *si se rechaza la hipótesis nula, en el contraste de ANOVA, eso significa que* ***no todas las medias son iguales****. Sin embargo, puede que* ***algunas*** *sí que sean iguales.*

Para decidir qué grupos tienen medias similares, descomponemos los niveles del factor en ***grupos homogéneos.***

PRUEBAS NO PARAMÉTRICAS

Contraste de Kruskal-Wallis

- Método no-paramétrico

- Util si fallan los requisitos del ANOVA (aunque inferior a ANOVA).

- Realiza un contraste sobre las medianas

H_0: $M_1 = M_2 = \ldots = M_n$
H_1: alguna M_i es distinta.

- Utiliza la noción de *rango*. La idea es ordenar de menor a mayor todos los datos (sin atender al nivel del factor del que provienen), asignar rangos, y comparar después los rangos medios correspondientes a los distintos niveles del factor.

¡Gracias!