EDEM

Máster en Data Analytics

Estadística con Python IV

Miguel Rua del Barrio

Comparación de múltiples medias

EDEM
Escuela de Empresarios

En la unidad anterior hemos visto test para comparar hasta 2 medias de dos grupos distintos.

A partir de ahora vamos a trabajar el caso de cuando tenemos más de dos grupos que comparar.

Ejemplo: Tiempo medio de recuperación del paciente

- **Según la enfermedad que padezcan (cáncer, bronquitis…)**
- **Por edad (mayor de 65, adulto, niño)**

Condiciones de aplicabilidad

EDEM
Escuela de Empresarios

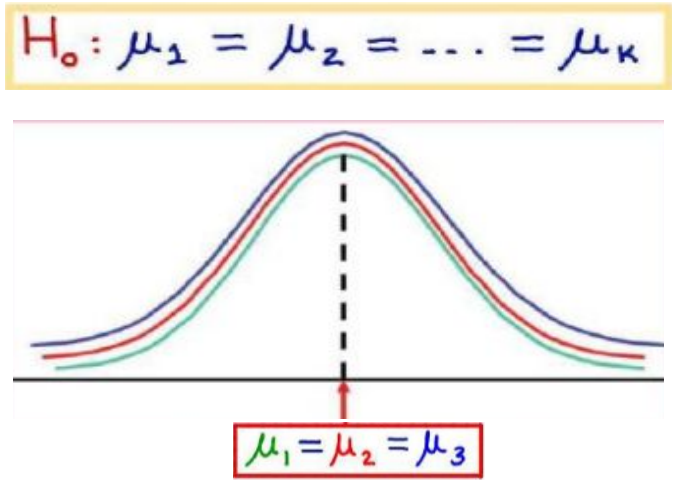
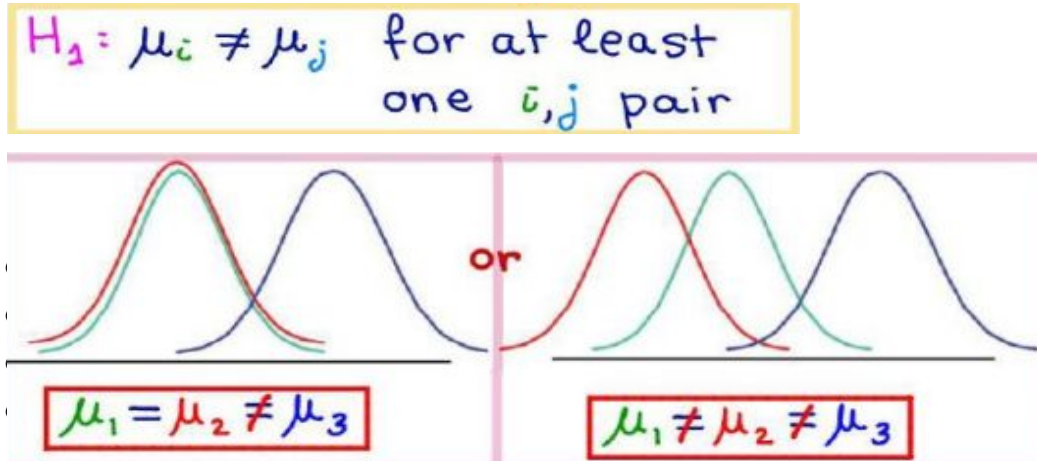
No podremos realizar estos análisis siempre que queramos, habrá unas condiciones de aplicabilidad:

1. **Poblaciones normalmente distribuidas en cada grupo.**
2. **Homocedasticidad de varianzas entre grupos.**
3. **Muestra aleatoria e independientes.**

Nos podremos ayudar de las técnicas vistas anteriormente para comprobar las condiciones de aplicabilidad.

Comparación de múltiples medias

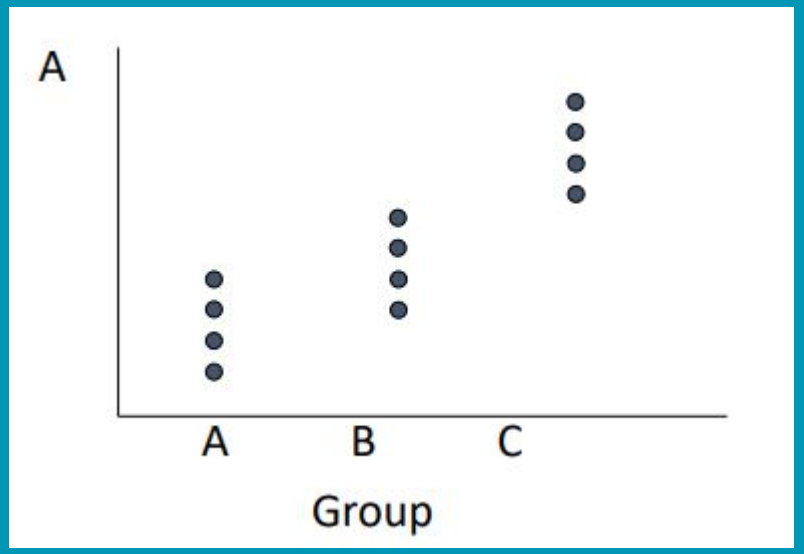
EDEM
Escuela de Empresarios

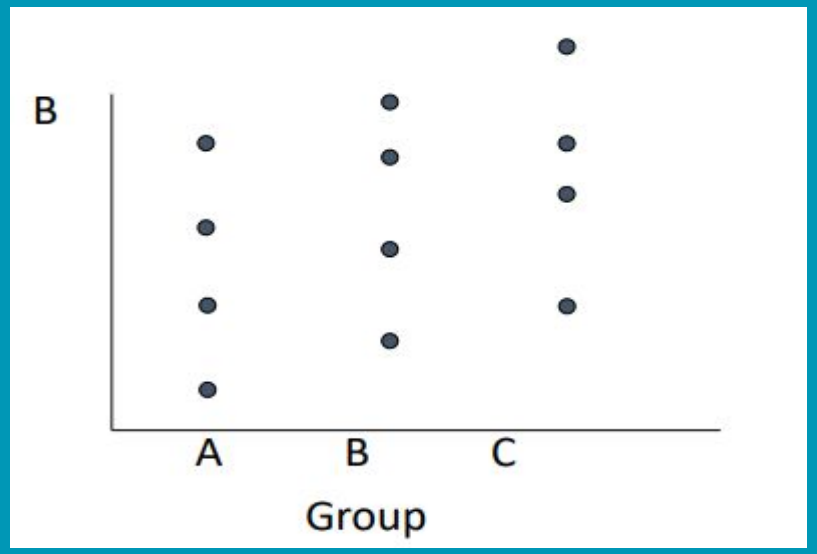
Hipótesis nula	Hipótesis alternativa
• **Todas las medias son iguales**  • **No hay diferencia entre las medias de los grupos**	• **Al menos una media es diferente**  • **Algunos pares de medias pueden ser iguales**

Variabilidad de los datos

EDEM
Escuela de Empresarios

- **La variabilidad es un factor clave para comprobar la igualdad de medias.**
- **En cada caso, las medias parecen diferentes, pero una gran variabilidad en el grupo B hace que no sea tan significativa esta diferencia.**

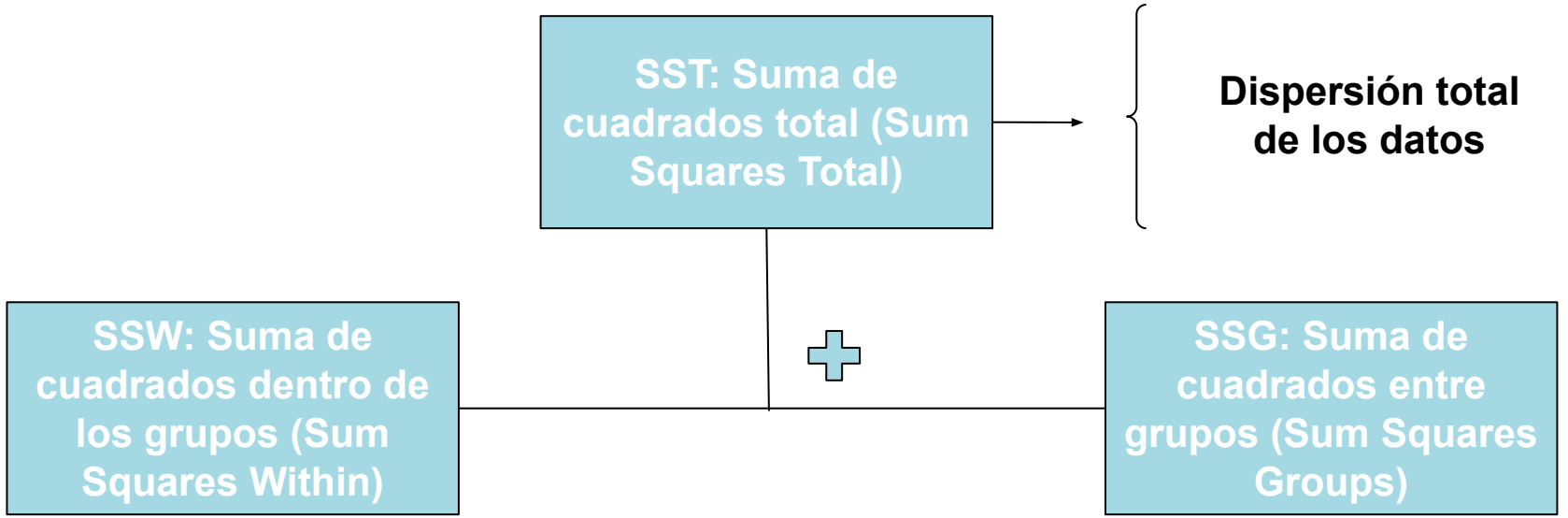




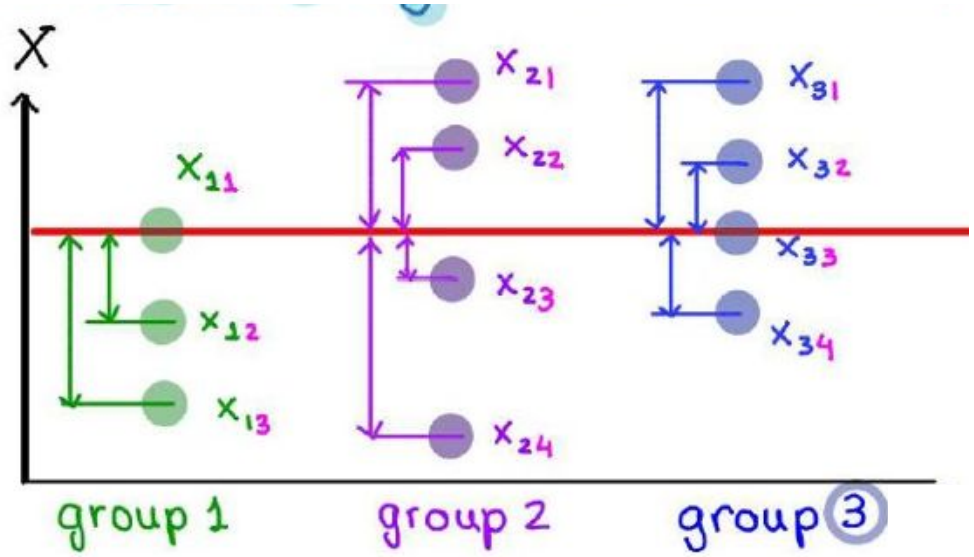
Comparación de múltiples medias

EDEM
Escuela de Empresarios

- **Enfoque: Estudiar variabilidad de los datos para realizar el test**



Variabilidad SST



$\bar{\bar{x}}$ = *media total de los datos*

n_i = *número de observaciones en el grupo i*

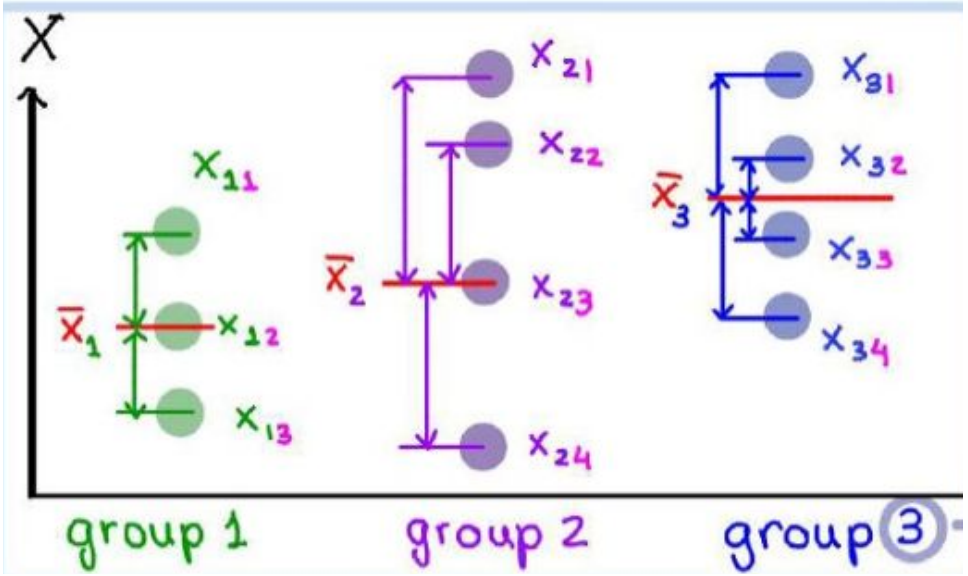
k = *número de grupos*

$$SST = \left[(x_{11} - \bar{\bar{x}})^2 + (x_{12} - \bar{\bar{x}})^2 + (x_{13} - \bar{\bar{x}})^2 + (x_{21} - \bar{\bar{x}})^2 + \ldots + (x_{24} - \bar{\bar{x}})^2 + (x_{31} - \bar{\bar{x}})^2 + \ldots + (x_{34} - \bar{\bar{x}})^2 \right] = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2$$

x_{ij}: jth observation from group i

Variabilidad SSW

EDEM
Escuela de Empresarios



$\bar{x}_i$ = *media total de los datos*

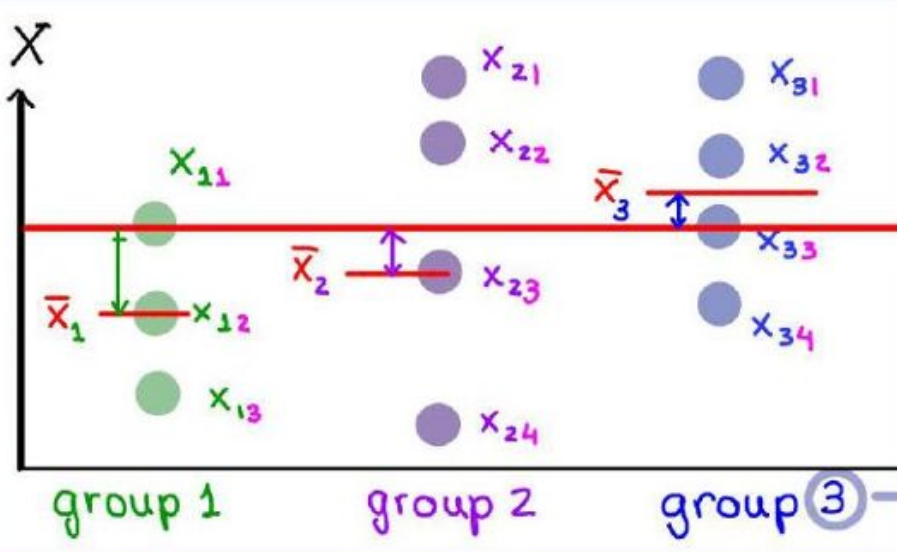
n_i = *número de observaciones en el grupo i*

k = *número de grupos*

$$SSW = \left[(x_{11} - \bar{x}_1)^2 + (x_{12} - \bar{x}_1)^2 + (x_{13} - \bar{x}_1)^2 + (x_{21} - \bar{x}_2)^2 + \ldots + (x_{24} - \bar{x}_2)^2 + (x_{31} - \bar{x}_3)^2 + \ldots + (x_{34} - \bar{x}_3)^2 \right] = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

x_{ij}: jth observation from group i

Variabilidad SSG



$\bar{x}_i$ = *media de cada grupo*

$\bar{x}$ = *media total de los datos*

n_i = *número de observaciones en el grupo i*

k = *número de grupos*

$$SSG = n_1 \cdot (\bar{x}_1 - \bar{\bar{x}})^2 + n_2 \cdot (\bar{x}_2 - \bar{\bar{x}})^2 + n_3 (\bar{x}_3 - \bar{\bar{x}})^2 = \sum_{i=1}^{k} n_i \cdot (\bar{x}_i - \bar{\bar{x}})^2$$

Test ANOVA

EDEM
Escuela de Empresarios

- **Hipótesis**

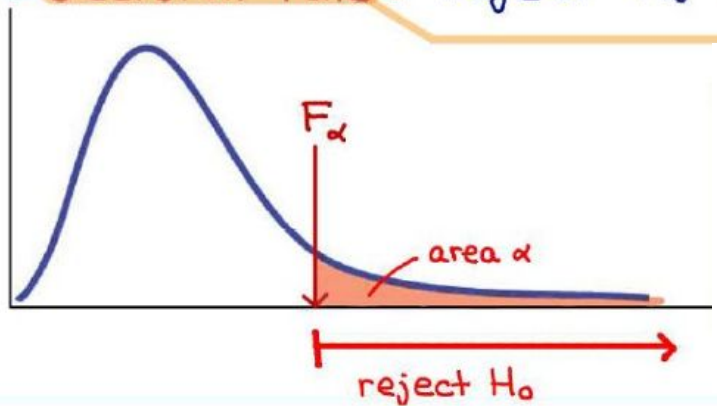
$$\begin{cases} H_0: \mu_1 = \mu_2 = \ldots = \mu_K \\ H_1: \text{at least 2 } \mu_i\text{'s are} \neq \end{cases}$$

- **Estadístico del test**

$$F = \frac{MSG}{MSW} = \frac{\text{between groups var}/(K-1)}{\text{within groups var}/(n-K)}$$

$$\begin{cases} df_1 = K-1 \quad \text{(typically small)} \\ df_2 = n-K \quad \text{(} \quad \text{"} \quad \text{large)} \end{cases}$$

* Decision rule: reject H_0 if $F > F_{critical} = F_{K-1,n-K,\alpha}$



ANOVA de un factor

EDEM
Escuela de Empresarios

- **Se suele utilizar la tabla ANOVA para representar el contraste**

Source of variation	SS	df	MS (variance)	F-ratio
Between groups	SSG	$k-1$	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSW}$
Within groups	SSW	$n-k$	$MSW = \frac{SSW}{n-k}$	
Total	SST	$n-1$		

ANOVA de un factor: ejemplo

EDEM
Escuela de Empresarios

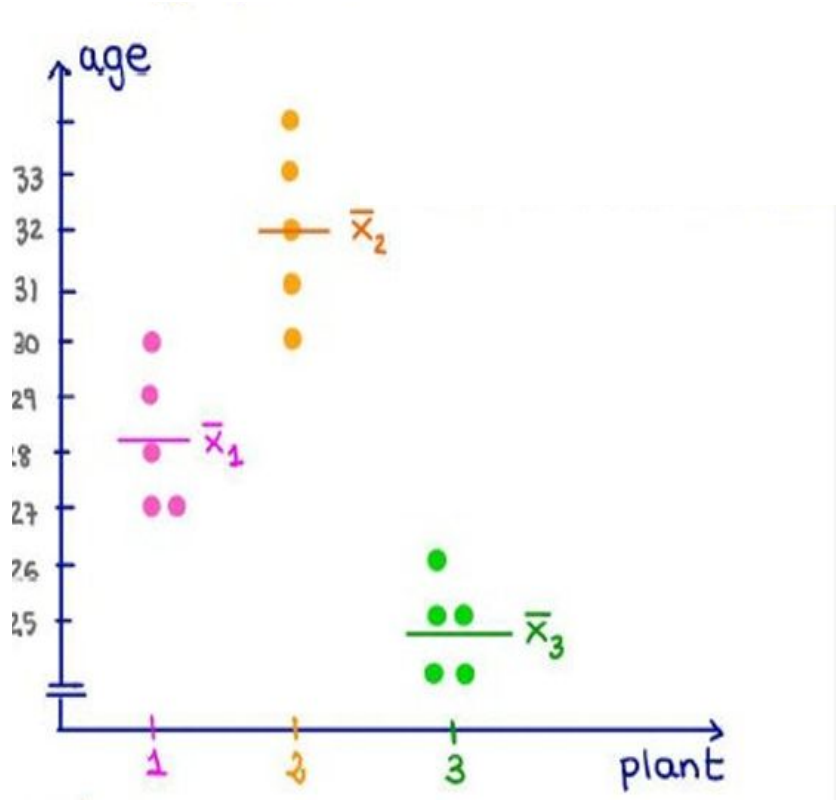
- **Se suele utilizar la tabla ANOVA para representar el contraste**

P-1	P- 2	P- 3
29	32	25
27	33	24
30	31	24
27	34	25
28	30	26

$\bar{x}_1 = 28.2$ $\bar{x}_2 = 32$ $\bar{x}_3 = 24.8$

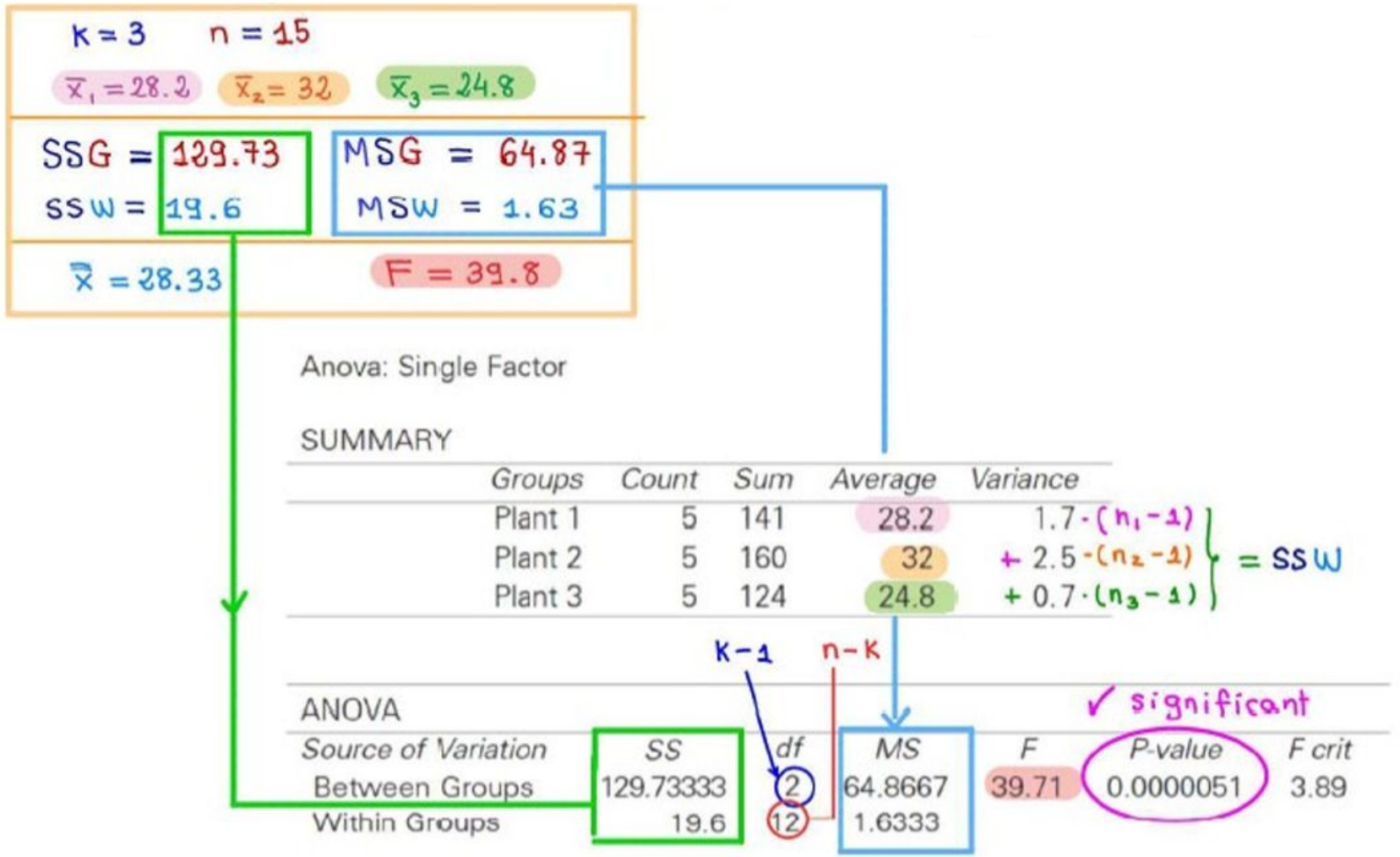
$k = 3$ $\bar{\bar{x}} = 28.33$

$n_1 = n_2 = n_3 = 5 \Rightarrow n = 15$



ANOVA de un factor: ejemplo

EDEM
Escuela de Empresarios



ANOVA de un factor: ejemplo

EDEM
Escuela de Empresarios

* Hypotheses: $\{ H_0 : \mu_1 = \mu_2 = \mu_3 \;;\; H_1 : \mu_i \text{ not all } = \}$

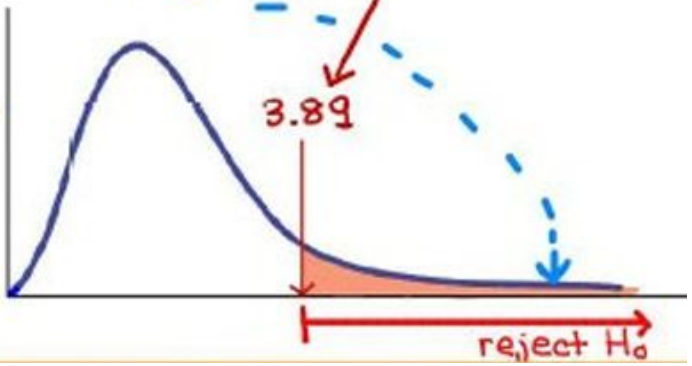
$\alpha = 0.05$

* Test statistic: $F = 39.71$

p-value for the F-test

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	129.73333	2	64.8667	39.71	0.0000051	3.89
Within Groups	19.6	12	1.6333			



* Decision rule: reject H_0 if $F > F_{2,12,0.05} = 3.89$

Decisión: como F ∈ a la región de rechazo (o p valor < α), rechazamos Ho

Conclusión: existen diferencias significativas en la edad media de los tres grupos

EDEM
Escuela de Empresarios

Parte práctica de la sesión:

1. Realizar ANOVA con Python y test ad hoc

Resumen

EDEM
Escuela de Empresarios

Durante este módulo hemos visto:

- **Nociones básicas de estadística como su nomenclatura (población, muestra..).**
- **Definición de las variables aleatorias y sus diferentes tipos.**
- **Descriptiva univariante y bivariante para cada tiempo de variable (gráficos y estadísticos)**
- **Distribuciones de probabilidad (definiciones, ejemplos)**
- **Contrastes de hipótesis (definición, concepto, ejemplos).**
- **ANOVA.**
- **Implementación de todos estos contenidos con Python.**
- **Realización de informes estadísticos con Google Colab.**

EDEM
Escuela de Empresarios

¡GRACIAS POR VUESTRA ATENCIÓN!

miguel.ruadelbarrio@ams-europe.com

linkedin.com/in/miguel-rua-del-barrio-5214661b5