



Instituto  
**CBOQ**

*Educamos diferente*

ANOVA

ANALYSIS OF VARIANCE

# DEFINICIÓN

- Es un grupo de modelos estadísticos y los procedimientos de estimación asociados a estos, utilizados para analizar la diferencia de medias entre grupos de muestras.
- El procedimiento fue desarrollado por Fisher y está basado en la ley total de varianza que enuncia que la varianza observada de una variable particular se puede separar en los componentes atribuibles a los diferentes orígenes de la variación.
- En resumen, ANOVA provee un test estadístico para saber si las medias de dos o más poblaciones son iguales, generalizando la prueba de T para dos medias.

# EJEMPLO DE MOTIVACIÓN TEÓRICA

- ANOVA tiene utilidad como herramienta exploratoria para explicar observaciones.
- Tomando como ejemplo un concurso de perros, los concursantes generalmente son adultos, pura raza. Un histograma de los pesos de los perros será bastante complejo. Para simplificar su análisis y poder predecir el peso de un perro es útil separarlos por un grupo de características.
- Podríamos imaginar distintas maneras de agruparlos, por ejemplo utilizar la interacción de dos características binarias para obtener cuatro grupos, joven-senior y pelo corto-pelo largo; la distribución de peso de los subgrupos tiene alta varianza, y media similar. Por lo tanto esta agrupación no permite explicar la variación en el peso de los perros



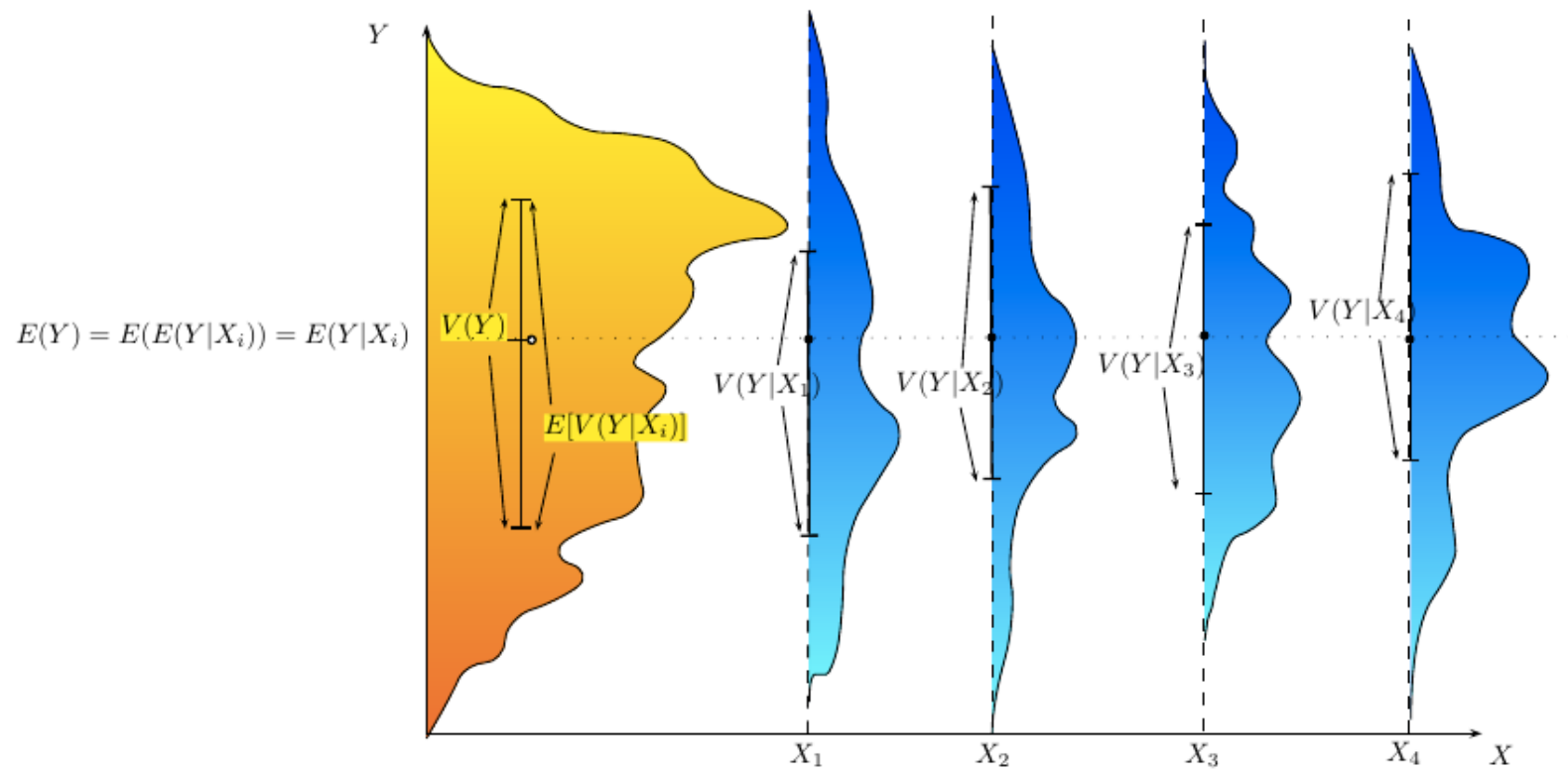


Figure 2: ANOVA : No fit

- Si agrupáramos los perros en base a las variables dicotómicas mascota-raza de trabajo y menos atlético-mas atlético obtendríamos una diferenciación un poco mas satisfactoria. Generalmente los perros mas pesados son los de disposición mas atlética y criados como razas de trabajo. Esto produce medias mas diferenciables, pero todavía hay cierta superposición de las distribuciones

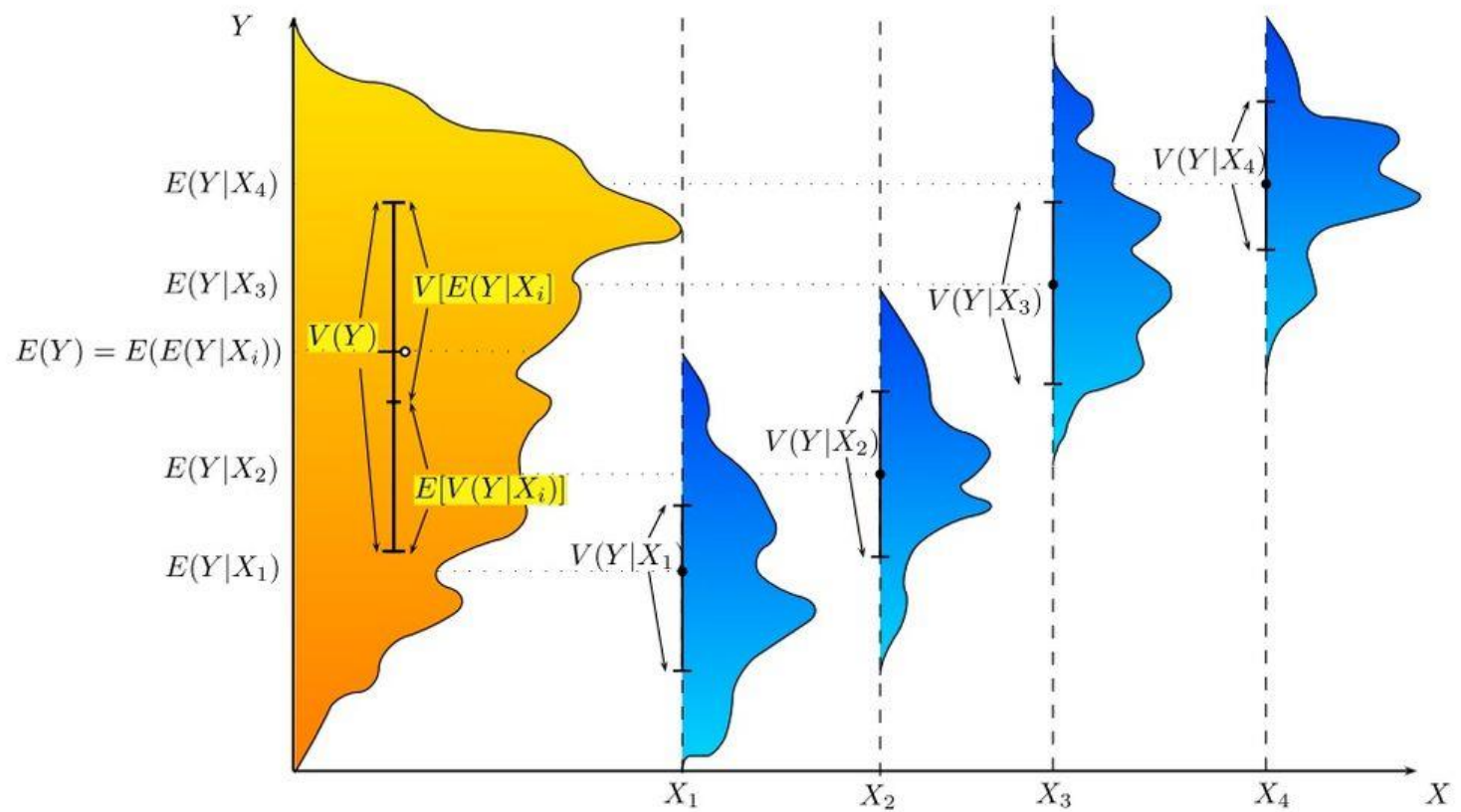


Figure 1: ANOVA : Fair fit.

- Generalmente, explicar la distribución de peso por la raza generalmente produce una mejor diferenciación. Todos los chihuahuas son pequeños y los San Bernardos gigantes.
- Aunque para esta explicación no es necesaria una herramienta estadística ya que es intuitivo por nuestra experiencia previa el ANOVA nos provee las herramientas formales para justificar estos análisis intuitivos y poder extenderlos a otros casos.



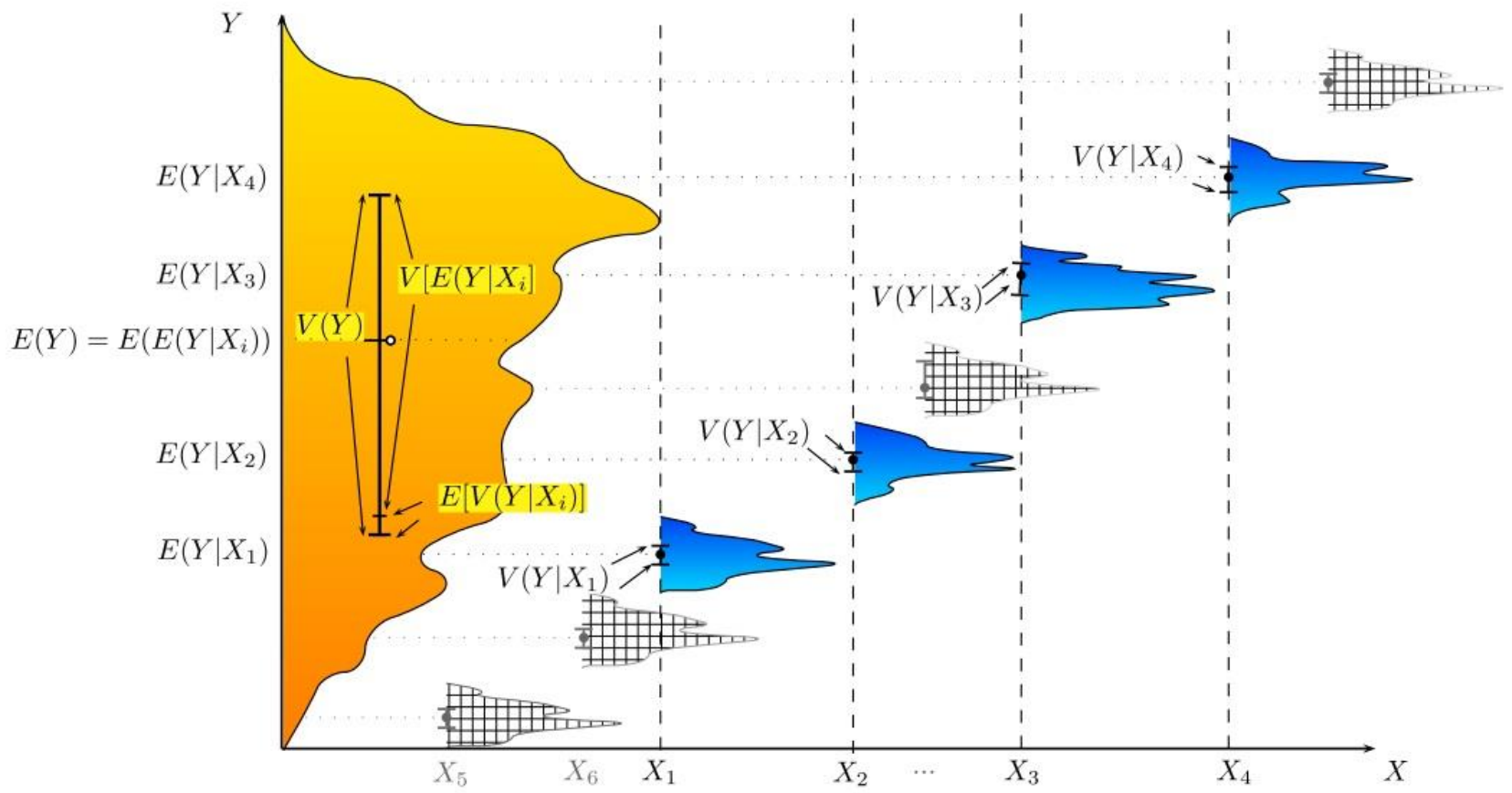


Figure 3: ANOVA : very good fit

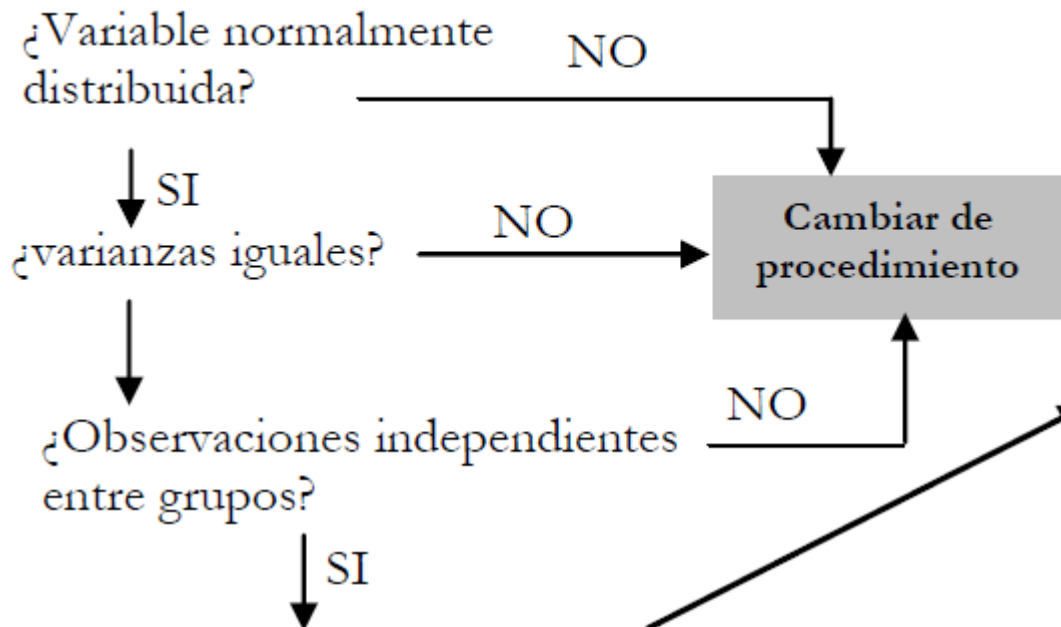


# CARACTERÍSTICA

- La herramienta ANOVA permite dividir el origen de la variación debida al azar (dentro de grupos) con la que es efectivamente causada por el factor de estudio (entre grupos).
- El procedimiento nos permite saber si la variabilidad entre grupos es significativamente diferente a la de dentro de los grupos.
- La hipótesis nula asume igualdad de medias y la alternativa por lo menos una media difiere.

# SUPUESTOS NECESARIOS

- Observaciones independientes
- Distribución de grupos normal
- Homocedasticidad (igualdad de varianzas entre grupos)



Fuente de Variación	SC	gl	MC	F
Entre grupos (tratamientos)	SC entre	a-1	MC entre	F <sub>calculado</sub>
Dentro de grupos (residual)	SC dentro	n-a	MC dentro	
<b>total</b>	SC total	n-1		

$$SC \text{ total} = \sum x^2 - (T^2/n)$$

$$SC \text{ entre} = \sum T_i^2/n_i - (T^2/n)$$

$$SC \text{ dentro} = SC \text{ total} - SC \text{ entre}$$

$$MC \text{ entre} = SC \text{ entre} / (a - 1)$$

$$MC \text{ dentro} = SC \text{ dentro} / (n - a)$$

$$F_c = MC \text{ entre} / MC \text{ dentro}; gl = (a - 1); (n - a)$$

**a** = número de grupos

**n<sub>i</sub>** = número de observaciones del i-ésimo grupo

**n** = total de observaciones

**T<sub>i</sub>** = Suma de observaciones del i-ésimo grupo

**T** = Suma total de observaciones



# EJERCICIO 1

En un experimento, se forman 4 grupos similares que son sometidos a distintos tratamientos. Se mide la variable de respuesta y se obtienen los siguientes resultados:

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	F calc
ENTRE			302	
DENTRO				
TOTAL	3213	22		

Completar la tabla, plantear las hipótesis y emitir conclusión con alfa de 5%

## EJERCICIO 2

La siguiente tabla muestra el contenido de nitrógeno, en miligramos, de plantas inoculadas con cultivos de diferentes cepas de *Rhizobium* (I-VI). El experimento se realizó utilizando un diseño aleatorio con grupos separados.

I	II	III	IV	V	VI
19.4	17.7	17.0	20.7	14.3	17.3
32.6	24.8	19.4	21.0	14.4	19.4
27.0	27.9	9.1	20.5	11.8	19.1
32.1	25.2	11.9	18.8	11.6	16.9
33.0	24.3	15.8	18.6	14.2	20.8

¿Las medias de los grupos difieren?

## EJERCICIO 4

Para probar si algún tratamiento experimental tenía efectos, en comparación con el placebo se aplicaron los 3 tratamientos y placebo a 4 grupos, obteniendo los siguientes resultados:

A	2	5	3	1	4	4
B	1	0	0	1	3	
C	6	4	5	5	6	2
D	8	7	6	5		

Comparar el efecto de los tratamientos con un 95% de confianza