



Mi Universidad

SUPER NOTA

Nombre del Alumno: ING.CHRISTIAN ESTÉVEZ HIDALGO

Nombre del tema

Parcial 1

Nombre de la Materia: SEMINARIO INTEGRADOR II

Nombre de la Licenciatura: MAESTRIA

Cuatrimestre

Tuxtla Gutiérrez, Chiapas. A 17 de octubre del 2022



ANÁLISIS DE LA VARIANZA CON UN FACTOR (ANOVA)

El análisis de la varianza permite contrastar la hipótesis nula de que las medias de K poblaciones ($K > 2$) son iguales, frente a la hipótesis alternativa de que por lo menos una de las poblaciones difiere de las demás en cuanto a su valor esperado. Este contraste es fundamental en el análisis de resultados experimentales, en los que interesa comparar los resultados de K 'tratamientos' o 'factores' con respecto a la variable dependiente o de interés.

El Anova requiere el cumplimiento los siguientes supuestos:

Las poblaciones (distribuciones de probabilidad de la variable dependiente correspondiente a cada factor) son normales.

Las K muestras sobre las que se aplican los tratamientos son independientes.

Las poblaciones tienen todas igual varianza (homoscedasticidad).

El ANOVA se basa en la descomposición de la variación total de los datos con respecto a la media global (SCT), que bajo el supuesto de que H_0 es cierta es una estimación de obtenida a partir de toda la información muestral, en dos partes:

Variación dentro de las muestras (SCD) o Intra-grupos, cuantifica la dispersión de los valores de cada muestra con respecto a sus correspondientes medias.

Variación entre muestras (SCE) o Inter-grupos, cuantifica la dispersión de las medias de las muestras con respecto a la media global.

Con los datos de la encuesta sobre transporte, Enctrans.sav, razonar si puede aceptarse que el tipo de transporte utilizado, Trans, influye sobre la variable tiempo.

Con la opción de menú **Gráficos > Barras de error > Simple** y con el botón **Definir** se selecciona como **Variable Tiempo** y en **Eje de categorías** la variable **Trans**; al aceptar se obtiene la siguiente representación gráfica

Como puede observarse, los puntos que representan a las medias de cada grupo aparecen dispersos a diferentes niveles; sobre todo la media del grupo definido por el factor **Tren**. El intervalo de confianza para la media correspondiente al grupo definido por el factor **Metro** está contenido dentro del intervalo correspondiente al grupo definido por el factor **Bus**, así como, el intervalo correspondiente al factor **Coche** está contenido dentro de los intervalos correspondientes definidos por los factores **Metro** y **Otros**. El gráfico, por tanto, parece sugerir no una única población sino tres poblaciones con distintas medias.

Para realizar el análisis de la varianza propiamente dicho la secuencia es **Analizar > Comparar medias > ANOVA de un factor**. En el cuadro de diálogo se selecciona **Tiempo** como variable **Dependiente** y **Trans** como **Factor**. Para contrastar la hipótesis de igualdad de varianzas se abre con el botón correspondiente el cuadro de diálogo **ANOVA de un factor: Opciones** y se activa **Homogeneidad de varianzas**. Si se desea

un análisis descriptivo del comportamiento de la variable dependiente dentro de cada grupo se activa también la opción Descriptivos. Al aceptar se obtienen los siguientes cuadros de resultados

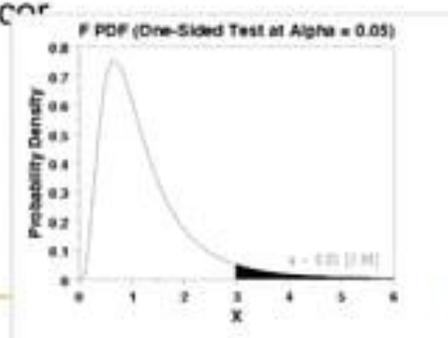
Este cuadro contiene un análisis descriptivo de la variable dependiente por grupos, así como, los límites superior e inferior para la media de cada grupo al 95% de confianza.

El estadístico de Levene toma un valor lo suficientemente pequeño para no rechazar la hipótesis de homocedasticidad a los niveles de significación habituales.

En el cuadro de resultados del ANOVA, el valor del estadístico de prueba, $F=6,450$, es significativamente distinto de 1 para cualquier nivel de significación y, por lo tanto, se rechaza la hipótesis nula de igualdad de medias y queda confirmada la primera impresión proporcionada por el gráfico de barras de error.

Pruebas de significación

- El análisis de varianza lleva a la realización de pruebas de significación estadística, usando la denominada distribución F de Snedecor



Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Sample	3	43.62	14.540	3.30	0.031
Error	36	158.47	4.402		
Total	39	202.09			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
2.09805	21.58%	15.05%	3.19%

Medias

Sample	N	Media	Desv.Est.	IC de 95%
1	10	11.203	1.995	(9.857, 12.548)
2	10	8.938	2.980	(7.592, 10.283)
3	10	10.683	1.102	(9.337, 12.028)
4	10	8.838	1.879	(7.492, 10.184)

Concepto

- El análisis de la varianza (ANOVA) es una colección de modelos estadísticos y sus procedimientos asociados, en el cual la varianza está particionada en ciertos componentes debidos a diferentes variables explicativas.
- El análisis de la varianza parte de los conceptos de regresión lineal.

Tabla 2 Expresiones para el cálculo del ANOVA de un factor (K indica el número de laboratorios y N el número total de resultados).

Fuente	Suma de cuadrados	Grados de libertad	Varianza	F_{cal}
Entre laboratorios	$SS_{lab} = \sum_{k=1}^K n_k (\bar{x}_k - \bar{\bar{x}})^2$	$K - 1$	$MS_{lab} = \frac{SS_{lab}}{K - 1}$	$F = \frac{MS_{lab}}{MS_R}$
Dentro de los laboratorios	$SS_R = \sum_{k=1}^K \sum_{j=1}^{n_k} (x_{kj} - \bar{x}_k)^2$	$N - K$	$MS_R = \frac{SS_R}{N - K}$	
Total	$SS_T = \sum_{k=1}^K \sum_{j=1}^{n_k} (x_{kj} - \bar{\bar{x}})^2$	$N - 1$	$MS_T = \frac{SS_T}{N - 1}$	

Análisis de Varianza (ANOVA): ¿qué es?

En estadística, encontramos el concepto de análisis de Varianza (ANOVA), que consiste en **una agrupación de modelos estadísticos y sus procedimientos asociados, donde la varianza está particionada en ciertos componentes**, debido a variables explicativas diversas. Si desglosamos sus siglas en inglés, ANOVA significa: ANalysis Of VAriance (Análisis de la varianza).

El Análisis de Varianza (ANOVA) es **un tipo de prueba paramétrica**. Esto quiere decir que deben cumplirse una serie de supuestos para aplicarla, y que el nivel de la variable de interés debe ser, como mínimo, cuantitativo (es decir, como mínimo de intervalo, por ejemplo el coeficiente intelectual, donde existe un 0 relativo).

Técnicas de análisis de varianza

Las primeras técnicas de análisis de varianza fueron desarrolladas en los años '20 y '30 por R.A. Fisher, un estadístico y genetista. Es por ello que el análisis de Varianza (ANOVA) **también se conoce como “Anova de Fisher” o “análisis de varianza de Fisher”**; esto también es debido al uso de la distribución F de Fisher (una distribución de probabilidad) como parte del contraste de hipótesis.

El análisis de varianza (ANOVA) **surge de los conceptos de regresión lineal**. La regresión lineal, en estadística, es un modelo matemático que se utiliza para aproximar la relación de dependencia entre una variable dependiente Y (por ejemplo la ansiedad), las variables independientes Xi (por ejemplo diferentes tratamientos) y un término aleatorio.

¿Qué es el análisis de covarianza?

Definición:

Es una prueba estadística que analiza la relación entre una variable dependiente y dos o más independientes, removiendo y controlando el efecto de al menos una de estas independientes

Se destacan **tres perspectivas** para el análisis de covarianza

- a) Perspectiva experimental
- b) Perspectiva de interés por la covariable
- c) Perspectiva de regresión

ANÁLISIS DE VARIANZA

SUPUESTOS

- 1.- Población: normal
- 2.- Muestreo y asignación: aleatorios
- 3.- Nivel de medición: intervalo o razón
- 4.- Varianza: σ^2
- 5.- Grupos independientes: Homogeneidad de varianzas (Homoscedasticidad)

Prueba robusta: relativamente insensible a la violación de los supuestos matemáticos

¿Qué es el análisis de varianza unidireccional ? (oneway)

Definición

Es una prueba estadística para analizar si más de dos grupos difieren significativamente entre sí en cuanto a sus medias y varianzas .

La prueba " t " es utilizada para dos grupos y el análisis de varianza unidireccional se usa para tres, cuatro o más grupos

Hipótesis a probar:

De diferencia entre dos o más grupos, se propone que los grupos difieren significativamente entre si y la nula propone que no difieren

18

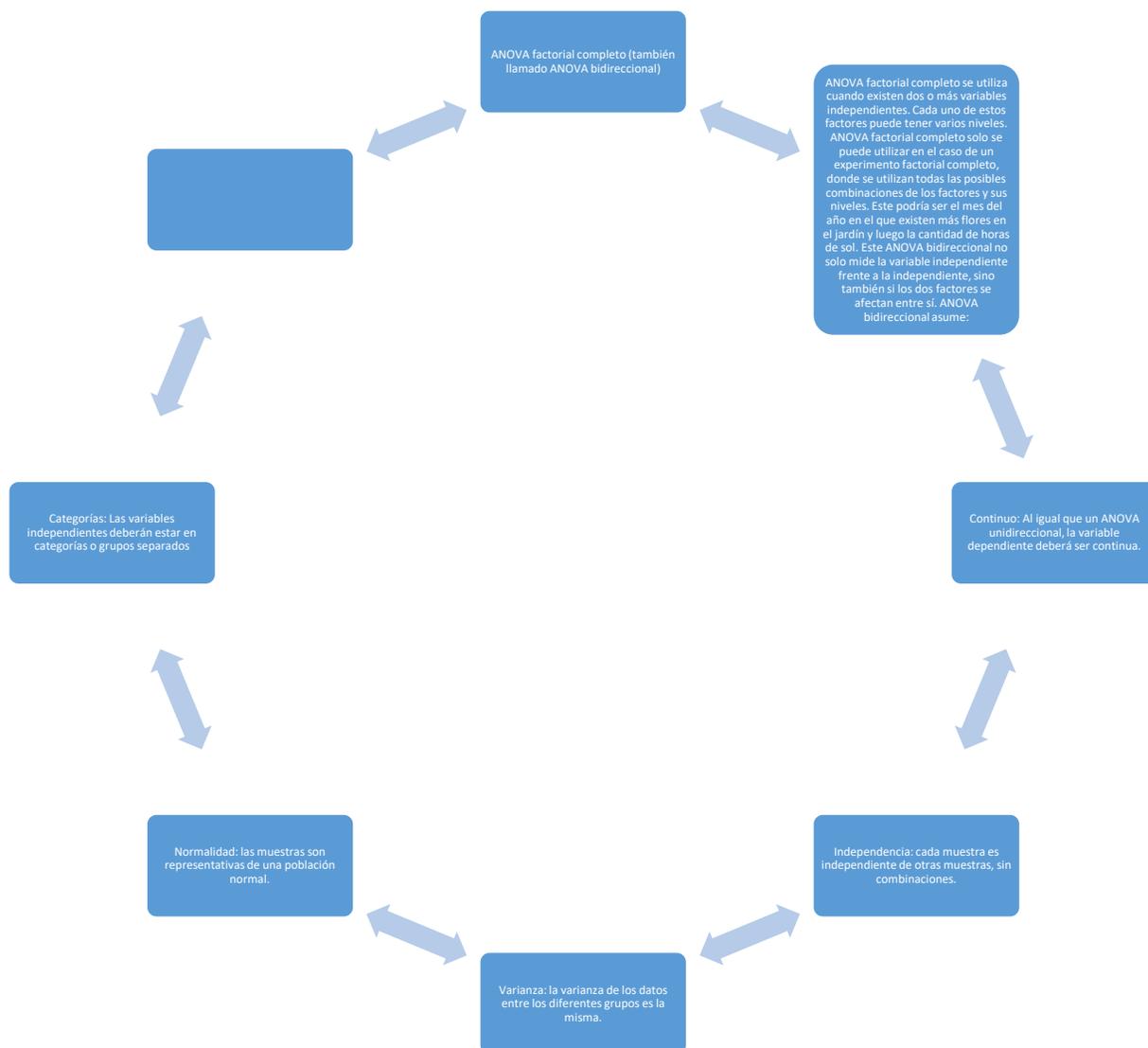
La varianza

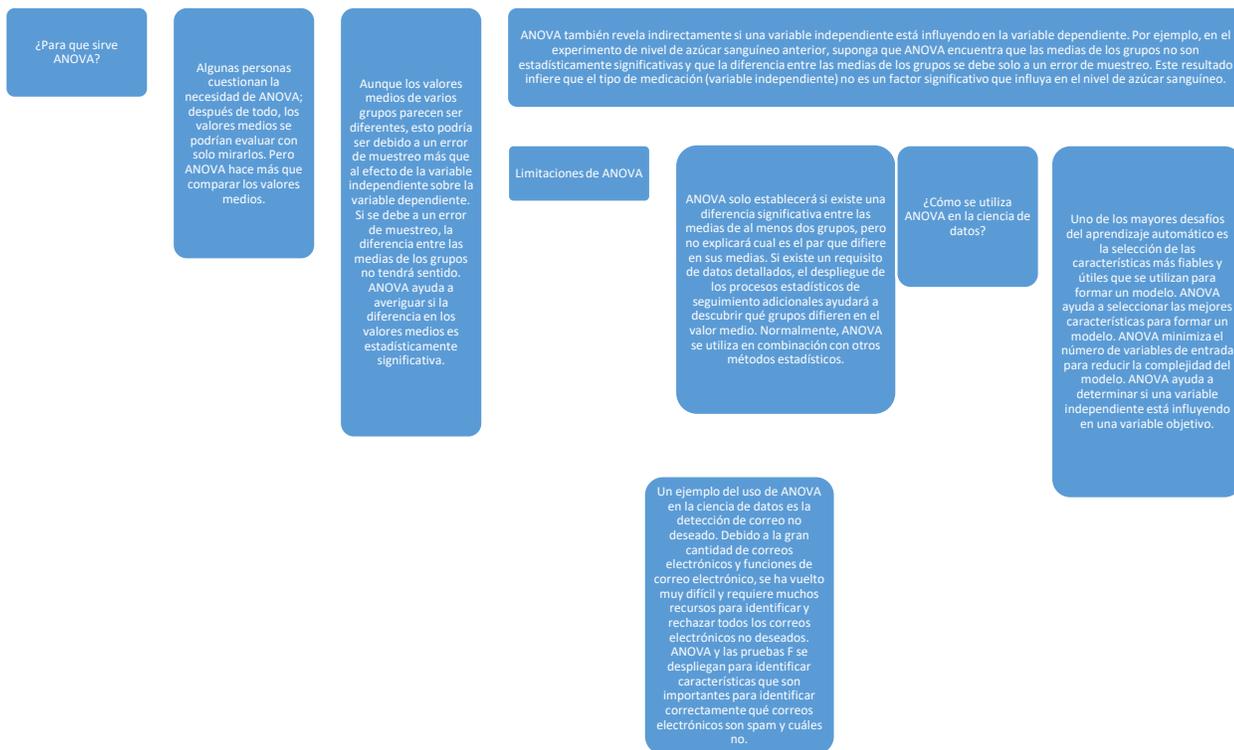
- Uno de los conceptos más importantes en el análisis estadístico y el control experimental de variables es la varianza.
- En principio, la varianza es una medida de variabilidad que da cuenta del grado de homogeneidad de un grupo de observaciones, la fórmula de cálculo es la siguiente:

$$\sigma_X^2 = \frac{\sum_{i=1}^N (X_i - \mu_X)^2}{N}$$

5/7/2012 La varianza y la inferencia estadística - Enrique Morosini 3







16							
17	ANÁLISIS DE VARIANZA						
18	<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
19	Entre grupos	1183,16667	2	591,5833333	6,0675214	0,021457405	4,256494729
20	Dentro de los grupos	877,5	9	97,5			
21							
22	Total	2060,66667	11				
23							

Valor F es mayor a valor crítico

$$V = \frac{\sum_{i=1}^{i=n} (X_i - \bar{X})^2}{n - 1}$$

ANOVA – Ejemplo:

- Media general
 $m = 40.53$
- Media por tratamiento:
 $m_A = 37, m_B = 41.3, m_C = 43.3$
- $SC_{total} = \sum (x_{iT} - 40.53)^2 \forall i, T.$
 29 Grados de libertad
- $SC_{tratamientos} = N \sum (m_T - m)^2$
 $N = n^\circ \text{ datos por tratamiento.}$
 2 grados de libertad.
- $SC_{residual} = \sum_T \sum_i (x_{iT} - m_T)^2$
 27 grados de libertad
 $SC_{total} = SC_{tratamiento} + SC_{residual}$
- $CM_{tratamiento} = SC_{tratamiento} / g_{tratamiento}$
- $CM_{residual} = SC_{residual} / g_{residual}$
- $F = CM_{tratamiento} / CM_{residual} = 0.532$
- Calcular $F_{2,27} (5\%) = 3.35 > 0.532$

	Factor estudiado		
	A	B	C
Resultados obtenidos	23	35	50
	28	36	43
	21	29	36
	27	40	34
	95	43	45
	41	49	52
	37	51	52
	30	28	43
	32	50	44
	36	52	34
Media	37	41.3	43.3

Comparación de Medias:

$\text{Var}(\hat{\phi}) = \sigma^2_{\hat{\phi}} = 0.66\sigma^2$

Un estimador de la $\sigma^2_{\hat{\phi}}$ se consigue sustituyendo σ^2 por el cuadrado medio del error (C.M.E.) obtenido de la tabla del análisis de varianza. (C.M.E. = 20.22)

$S^2_{\hat{\phi}} = (0.66)(20.22) = 13.34$

Probar $T_1 = T_2$ en oposición a $T_1 \neq T_2$ es equivalente a probar $H_0: Q = 0$ en oposición a $Q \neq 0$, donde $Q = T_1 - T_2$ por lo tanto la estadística adecuada es:

$$t_o = \frac{\hat{Q} - 0}{S_{\hat{Q}}}$$

11 Análisis de varianza de un factor

12

13 RESUMEN

14	Grupos	Cuenta	Suma	Promedio	Varianza
15	Matemáticas	7	603	86,14285714	75,8095238
16	Historia	4	264	66	46,6666667
17	Ciencias Naturales	5	420	84	197,5

18

19

20 ANÁLISIS DE VARIANZA

21	Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor critico para F
22	Entre grupos	1125,080357	2	562,5401786	5,2807052	0,020956726	3,805565253
23	Dentro de los grupos	1384,857143	13	106,5274725			
24							
25	Total	2509,9375	15				

26

REFERENCIAS

Pardo, A., y Ruiz, M.A. (2005). Análisis de datos con SPSS 13 Base. Ma-drid:

McGraw-Hill.Pardo, A., y San Martín, R. (1998). Análisis de datos en Psicología I

León, O.G., y Montero, I. (2001). Cómo explicar el concepto de interacción sin estadística: análisis gráfico de todos los casos posibles en un diseño 2×2. Psicothema, 13, 159-165I(2ªed.). Madrid: Pirámide