MODULO IV
ESTADÍSTICA EXPERIMENTAL

Probabilidad y Estadística

Ingeniería Industrial

FCA-UNNE

2020

Contenido

Tema 10: Análisis de la Variancia. Concepto. Fundamento matemático del análisis. Fuentes de variación. Grados de libertad. Sumas de cuadrados. Cuadrados medios. Razón de variancias para estimación de la significación.

¿CÓMO ANALIZAR LOS DATOS PROVENIENTES DE UN EXPERIMENTO?

El Análisis de la Variancia es una técnica estadística desarrollada por R. Fisher donde se particiona (descompone) la variabilidad total exhibida por un conjunto de datos en la suma de variabilidades parciales atribuibles a diferentes factores de variabilidad conocidos mas una variabilidad desconocida.

Vamos a ver a través de una situación cotidiana, como se piensa en la descomposición o partición la variabilidad total.

El cumpleaños de la abuela

La abuela Graciela cumplió 84 años y para festejarlo invitó a almorzar a sus nietos:

MARÍA, JUAN, JOSÉ y ESTELA, que están en el colegio de nivel medio.

ESTEBAN, EMILIO, ROSA y SARA, que están estudiando en la universidad.

SUSANA, RAÚL, ROQUE y PEDRO, que ya han terminado la universidad.

Después del almuerzo decidió repartir entre sus nietos, un premio que había ganado total ($840) y decide dividir el total en partes iguales para repartir.

NIETO	PESOS
MARÍA	70
JUAN	70
JOSÉ	70
ESTELA	70
ESTEBAN	70
EMILIO	70
ROSA	70
SARA	70
SUSANA	70
RAÚL	70
ROQUE	70
PEDRO	70
TOTAL	840

Sin embargo el abuelo interviene y dice que las necesidades económicas de los nietos son mayores para los mas grandes y propone cambios.

NIETO	PESOS	CAMBIOS Abuelo	CADA NIETO Recibe
MARÍA	70	-25	45
JUAN	70	-25	45
JOSÉ	70	-25	45
ESTELA	70	-25	45
ESTEBAN	70	-10	60
EMILIO	70	-10	60
ROSA	70	-10	60
SARA	70	-10	60
SUSANA	70	+35	105
RAÚL	70	+35	105
ROQUE	70	+35	105
PEDRO	70	+45	105
TOTAL	840	0	840

Luego de recibir el dinero, los nietos deciden jugar a las cartas agrupándose por nivel, juegan entre ellos los de nivel medio, los que están en la universidad y los que ya terminaron. Al final cada uno termina con ganancias, deudas o cómo llego.

NIETO	PESOS	CAMBIOS Abuelo	CADA NIETO Recibe	CAMBIOS Azar
MARÍA	70	-25	45	-30
JUAN	70	-25	45	+25
JOSÉ	70	-25	45	+15
ESTELA	70	-25	45	-10
ESTEBAN	70	-10	60	0
EMILIO	70	-10	60	+35
ROSA	70	-10	60	-40
SARA	70	-10	60	+5
SUSANA	70	+35	105	+15
RAÚL	70	+35	105	+20
ROQUE	70	+35	105	-30
PEDRO	70	+45	105	-5
TOTAL	840	0	840	0

¿Cómo termina el almuerzo y se van a sus casas?

NIETO	PESOS	CAMBIOS Abuelo	CAMBIOS Azar	RESULTADO
MARÍA	70	-25	-30	15
JUAN	70	-25	+25	70
JOSÉ	70	-25	+15	60
ESTELA	70	-25	-10	35
ESTEBAN	70	-10	0	60
EMILIO	70	-10	+35	95
ROSA	70	-10	-40	20
SARA	70	-10	+5	65
SUSANA	70	+35	+15	120
RAÚL	70	+35	+20	125
ROQUE	70	+35	-30	75
PEDRO	70	+45	-5	100
TOTAL	**840**	**0**	**0**	**840**

Raúl se va con 125 pesos, que se debe a lo que recibió de su abuela mas un poco de suerte. Sin embargo a Maria no le favoreció el azar y además recibió menos que Raúl.

Modelo Estadístico

A través de ese ejemplo se puede visualizar como expresar un dato (Y) a través de la suma de valores, mas abajo se presenta un modelo estadístico, el cual es necesario para entender como se particiona una varianza total.

UNIDAD DE ANALISIS.	μ	τ_i	ξ_{ij}	y_{ij}
MARÍA	70	-25	-30	15
JUAN	70	-25	+25	70
JOSÉ	70	-25	+15	60
ESTELA	70	-25	-10	35
ESTEBAN	70	-10	0	60
EMILIO	70	-10	+35	95
ROSA	70	-10	-40	20
SARA	70	-10	+5	65
SUSANA	70	+35	+15	120
RAÚL	70	+35	+20	125
ROQUE	70	+35	-30	75

ANÁLISIS DE LA VARIANCIA

El Análisis de la Variancia es una técnica estadística desarrollada por R. Fisher que permite particionar la variabilidad total exhibida por un conjunto de individuos provenientes de una muestra o población (σ_T^2), en una serie de variabilidades parciales ($\sigma_1^2 + \sigma_2^2 + ...$), atribuibles a diferentes fuentes o factores de variabilidad conocidos o sospechados y una variabilidad residual, no atribuible a causas conocidas, denominada variabilidad aleatoria o error experimental (σ_ξ^2).

$$\sigma_T^2 = \sigma_1^2 + \sigma_2^2 + ... + \sigma_\xi^2$$

Ejemplo

Supongamos que tenemos datos (y) que provienen de R tratamientos, nos interesa comparar los tratamientos , además cada tratamiento esta repetido r veces .

		Repeticiones					
		1	2	3	.	r	Prome-dios
Tratamientos	1	Y_{11}	Y_{12}	Y_{13}	.	Y_{1r}	$\bar{y}_1$
	2	Y_{21}	Y_{22}	Y_{23}	.	Y_{2r}	$\bar{y}_2$
	.	.	.	.	.	.	
	.	.	.	.	.	.	$\bar{y}_2$
	.	.	.	.	.	.	
	R	Y_{R1}	Y_{R2}	Y_{R3}	.	Y_{Rr}	$\bar{y}_R$

H0: $\mu_1 = \mu_2 = \mu_3 = \cdots . = \mu_R = \mu$ H1: *Al menos un* $\mu_i \neq \mu$

No es suficiente ver los promedios y decir que ellos difieren, es necesario una prueba.
La prueba consistirá comparar varianzas.

Recordemos la fórmula de la Varianza

$$S^2 = \frac{\sum_{i=1}^{n} d_i}{n-1} = \frac{\sum_{i=1}^{n} (y_{ij} - \bar{y})^2}{n-1}$$

Al numerador lo llamamos Suma de Cuadrados (S.C)

Al denominador lo llamamos Grados de Libertad (G.L)

Para poder comparar promedios de tratamientos vamos a comparar Varianzas, entonces vamos a particionar la varianza total.

ANOVA

Si se consideran todos los desvíos de todas las observaciones de todas las muestras, se tendrá que:

$$\sum_{i=1}^{R} \sum_{j=1}^{r} \left(y_{ij} - y\right)^2 = \left[\sum_{i=1}^{R} \sum_{j=1}^{r} \left(y_{ij} - y_i\right)^2 \right] + \left[r \sum_{j=1}^{r} \left(y_i - y\right)^2 \right]$$

El miembro de la izquierda corresponde a las Suma de Cuadrados Total (SCT), que como se puede ver está integrada por dos componentes: una Suma de Cuadrados Dentro de Tratamientos (SCDTrat) y una Suma de Cuadrados Entre Tratamientos (SCETrat).

Demostración de la partición de la suma de cuadrados

Si se considera la diferencia o desvío de una observación cualquiera respecto de la media general, y se opera, se tiene:

$$\left(y_{ij}-y\right)^2=\left[\left(y_{ij}-y_i\right)+\left(y_i-y\right)\right]^2=\left(y_{ij}-y_i\right)^2+\left(y_i-y\right)^2+2\left(y_{ij}-y_i\right)\left(y_i-y\right)$$

Si se consideran los desvíos de todas las observaciones dentro de una de las muestras:

$$\sum_{j=1}^{r}\left(y_{ij}-y\right)^2=\sum_{j=1}^{r}\left(y_{ij}-y_i\right)^2+\sum_{j=1}^{r}\left(y_i-y\right)^2+2\sum_{j=1}^{r}\left(y_{ij}-y_i\right)\left(y_i-y\right)$$

Como dentro de una muestra $\left(y_i-y\right)$ es constante y $\sum_{j=1}^{r}\left(y_{ij}-y_i\right)=0$, se tiene que:

$$\sum_{j=1}^{r}\left(y_{ij}-y\right)^2=\sum_{j=1}^{r}\left(y_{ij}-y_i\right)^2+\sum_{j=1}^{r}\left(y_i-y\right)^2$$

Esas sumas de cuadrados se pondrán en un cuadro llamado ANOVA

Estas son las particiones de las Sumas de Cuadrados, para llegar a las particiones de las variancias (Cuadrados Medios), es necesario particionar los Grados de Libertad y hallar los cocientes entre Sumas de Cuadrados y Grados de Libertad, para lo cual se resume la información en el siguiente Cuadro:

Cuadro de Análisis de la Variancia

Factores de Variación F.V.	Grados de Libertad G.L.	Sumas de Cuadrados S.C.	Cuadrados Medios C.M.	F_c
Entre Muestras (Tratamientos)	*R-1*	*SCEntre Trat*	$\frac{SCETrat}{GLETrat}$	$\frac{CMETrat}{CMDTrat}$
Dentro de Muestras (Error Experimental)	*R(r-1)*	*SCDentro Trat*	$\frac{SCDTrat}{GLDTrat}$	
Total	*R.r-1*			

ANOVA

La columna de los Cuadrados Medios contiene las variancias atribuibles a las diferentes factores de variación. Allí se puede observar la importancia relativa de la porción de variabilidad con que contribuye cada fuente a la variabilidad total.

Se incluye en una última columna una prueba de razón de variancias (F), para probar la homogeneidad entre la variancia debida a causa conocida (Entre Muestras o Tratamientos) y a causas aleatorias o desconocidas (Dentro de Muestras o Error Experimental).

Que estas variancias sean homogéneas indica que la variabilidad debida a causas conocidas (muestras) no se diferencia de la variabilidad debida a causas desconocidas, azar o error experimental.

ANOVA

Estas son las particiones de las Sumas de Cuadrados, para llegar a las particiones de las variancias (Cuadrados Medios), es necesario particionar los Grados de Libertad y hallar los cocientes entre Sumas de Cuadrados y Grados de Libertad, para lo cual se resume la información en el siguiente Cuadro:

ANOVA

Cuadro de Análisis de la Variancia

Factores de Variación F.V.	Grados de Libertad G.L.	Sumas de Cuadrados S.C.	Cuadrados Medios C.M.	F_c
Entre Muestras (Tratamientos)	R-1	SCETrat	$\frac{SCDTrat}{GLDTrat}$	$\frac{CMETrat}{CMDTrat}$
Dentro de Muestras (Error Experimental)	R(r-1)	SCDTrat	$\frac{SCETrat}{GLETrat}$	
Total	R.r-1			

ANOVA

La columna de los Cuadrados Medios contiene las variancias atribuibles a las diferentes fuentes o factores de variación. Allí se puede observar la importancia relativa de la porción de variabilidad con que contribuye cada fuente a la variabilidad total.

Se incluye en una última columna una prueba de razón de variancias (F), para probar la homogeneidad entre la variancia debida a causa conocida (Entre Muestras o Tratamientos) y a causas aleatorias o desconocidas (Dentro de Muestras o Error Experimental).

Que estas variancias sean homogéneas indica que la variabilidad debida a causas conocidas (muestras) no se diferencia de la variabilidad debida a causas desconocidas, azar o error experimental.