ANOVA de 1 Factor: Otros Detalles

Clase 4

Nicolás Mejía M.
n.mejia10@uniandes.edu.co

Probabilidad y Estadística II
Departamento de Ingeniería Industrial
Universidad de Los Andes, Bogotá, Colombia

2020-19

Outline

Notación

Al organizar la información en una tabla:

	Factor A				
Obs (j)	Nivel 1	...	Nivel i	...	Nivel a
1	Y_{11}	...	Y_{i1}	...	Y_{a1}
2	Y_{12}	...	Y_{i2}	...	Y_{a2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	Y_{1j}	...	Y_{ij}	...	Y_{aj}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	Y_{1n_1}	...	Y_{in_i}	...	Y_{an_a}
Promedios columna	$\bar{Y}_{1.}$	...	$\bar{Y}_{i.}$	...	$\bar{Y}_{a.}$
Desv. Est.	S_1	...	S_i	...	S_a

Prueba ANOVA

$$H_0: \text{ El Factor NO influye sobre } Y \Leftrightarrow \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1: \text{ El Factor SI influye sobre } Y \Leftrightarrow \text{Algún par } \mu_i \neq \mu_j$$

Fuente	**SS**	**gl**	**MS**	**F**
Factor A	$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$a-1$	$SSA/(a-1)$	MSA/MSE
Error	$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2$	$N-a$	$SSE/(N-a)$	-
Total	$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2$	$N-1$	$SST/(N-1)$	-

Donde $SST = SSA + SSE$ y $gl_T = gl_A + gl_e$

La región de rechazo es:

$$RHO \Leftrightarrow F_{1-\alpha,a-1,N-a}$$

Outline

Un buen estimador

Con base a lo que se ha visto, en nuestro contexto de diseño de experimentos debería ser claro que el estimador muestral de la varianza usual S^2_{Total}, en nuestra notación el MST, no es un buen estimador de la varianza σ^2.

Esto es ya que, desde el punto de vista de probabilidad, la varianza hace referencia a la desviación inherente dada por el fenómeno aleatorio.

Si existe un factor significativo, ya se sabe que de toda la variación observada, hay una fracción importante que no es producto de la aleatoriedad, por ende hay un problema.

Bajo esa lógica, resulta claro que un mejor estimador de la varianza σ^2, sería aquel dado por el componente del ERROR.

Matemáticamente

Esto se puede demostrar matemáticamente. Especificamente, cuando H_0 es falsa se tiene que:

$$E\left(MSE\right) = \sigma^2$$

$$E\left(MSA\right) = \sigma^2 + \frac{\sum_{i=1}^{a} n_i \left(\mu_i - \mu\right)^2}{a-1}$$

$$E\left(MST\right) = \sigma^2 + \frac{\sum_{i=1}^{a} n_i \left(\mu_i - \mu\right)^2}{N-1}$$

Donde μ_i es la media del nivel i y $\mu = \left(a^{-1} \sum_{i=1}^{a} \mu_i\right)$ es la media total.

El mejor estimador

El cuadrado medio del error (MSE) es el único estimador insesgado de la varianza en todos los casos.

Si H_0 es cierta	Si H_1 es cierta (ó H_0 es falsa)
$E(MSE) = \sigma^2$	$E(MSE) = \sigma^2$
$E(MST) = \sigma^2$	$E(MST) > \sigma^2$
$E(MSA) = \sigma^2$	$E(MSA) > \sigma^2$

Así, de ahora en adelante el MSE será el mejor estimador de la varianza

$$\hat{\sigma}^2 = MSE$$

¡Tenganlo SUPER presente!

Implicaciones

¿En donde se utiliza el estimador de la varianza? En pruebas de medias e intervalos de confianza. Eso significa que los estadísticos cambian, en lugar de usar el estimador que allí se definía, se utilizará el MSE y sus grados de libertad.

Recuerden de Proba 1 que:

$$t = \frac{\bar{Y}_{..} - \mu}{\sqrt{S^2/N}} \sim t_{N-1}$$

Ya que sabemos que el MSE es el mejor estimador de la varianza el estadístico anterior queda:

$$t = \frac{\bar{Y}_{..} - \mu}{\sqrt{MSE/N}} \sim t_{N-a}$$

P-Valor

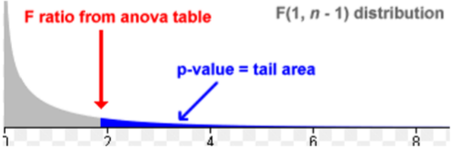
Otra forma de contrastar pruebas de hipótesis es mediante el uso del P-valor. El P-valor se define informalmente como "la probabilidad de que ocurra un fenómeno igual o más raro al que se observó, si la hipótesis nula fuera verdad".

Por eso, cuando la probabilidad de que el evento que observamos suceda es muy pequeña, pensamos realmente que la hipótesis original es dudosa.

Por tanto, el p-valor es una medida de la fuerza de la evidencia en sus datos en contra de H_0. Mientras más pequeño, más fuerte será la evidencia de la muestra para rechazar H_0.

P-Valor

Numericamente esta probabilidad esta dada por:



En este sentido ¿Cúal es la formula para hallar el P-valor de la prueba F en el diseño de experimentos?

$$PValor = P(F_{a-1,N-a} > F_{calculada})$$

Regla del P-valor

Si el P-valor es menor a la significancia de la prueba (α) se rechaza la hipótesis nula (H_0). Por el contrario si el P-valor es mayor a la significancia NO se reachaza la hipótesis nula.